10049137 , 020102

JC10 Rec'd PCT/PTO 07 FEB 2002

FORM PTO-1390 (REV. 12-2001) U S DEPARTMENT OF COMMERCE PATENT AND TRADEMARK OFFICE

TRANSMITTAL LETTER TO THE UNITED STATES DESIGNATED/ELECTED OFFICE (DO/EO/US) CONCERNING A FILING UNDER 35 U.S.C. 371

ATTORNEY 'S DOCKET NUMBER: 7024-509

U S APPLICATION NO (If known, see 37 CFR 1 5): 10/049137

INTERNATIONAL APPLICATION NO.	INTERNATIONAL FILING DATE	PRIORITY DATE CLAIMED
PCT/US00/22725	18 August 2000 (18.08.00)	20 August 1999 (20.08.99)

TITLE OF INVENTION
METHODS AND COMPOSITIONS FOR REGULATING DEVELOPMENTAL IDENTITY

APPLICANT(S) FOR DO/EO/US
Joseph P. OGAS; Chris R. SOMERVILLE

Applicant herewith submits to the United States Designated/Elected Office (DO/EO/US) the following items and other information:

1. [X] This is a **FIRST** submission of items concerning a filing under 35 U.S.C. 371.
2. [ ] This is a **SECOND** or **SUBSEQUENT** submission of items concerning a filing under 35 U.S.C. 371.
3. [ ] This is an express request to begin national examination procedures (35 U.S.C. 371(f)). The submission must include items (5), (6), (9) and (21) indicated below.
4. [ ] The US has been elected by the expiration of 19 months from the priority date (Article 31).
5. [X] A copy of the International Application as filed (35 U.S.C. 371(c)(2))
   a. [ ] is attached hereto (required only if not communicated by the International Bureau).
   b. [ ] has been communicated by the International Bureau.
   c. [X] is not required, as the application was filed in the United States Receiving Office (RO/US).
6. [ ] An English language translation of the International Application as filed (35 U.S.C. 371(c)(2)).
   a. [ ] is attached hereto.
   b. [ ] has been previously submitted under 35 U.S.C. 154(d)(4).
7. [X] Amendments to the claims of the International Aplication under PCT Article 19 (35 U.S.C. 371(c)(3))
   a. [ ] are attached hereto (required only if not communicated by the International Bureau).
   b. [X] have been communicated by the International Bureau.
   c. [ ] have not been made; however, the time limit for making such amendments has NOT expired.
   d. [ ] have not been made and will not be made.
8. [ ] An English language translation of the amendments to the claims under PCT Article 19 (35 U.S.C. 371 (c)(3)).
9. [X] An oath or declaration of the inventor(s) (35 U.S.C. 371(c)(4)). (unsigned)
10. [ ] An English lanugage translation of the annexes of the International Preliminary Examination Report under PCT Article 36 (35 U.S.C. 371(c)(5)).

Items 11 to 20 below concern document(s) or information included:

11. [ ] An Information Disclosure Statement under 37 CFR 1.97 and 1.98.
12. [ ] An assignment document for recording. A separate cover sheet in compliance with 37 CFR 3 28 and 3.31 is included.
13. [ ] A FIRST preliminary amendment.
14. [ ] A SECOND or SUBSEQUENT preliminary amendment.
15. [ ] A substitute specification.
16. [ ] A change of power of attorney and/or address letter.
17. [X] A computer-readable form of the sequence listing in accordance with PCT Rule 13ter.2 and 35 U S.C. 1 821 - 1.825
18. [ ] A second copy of the published international application under 35 U.S.C. 154(d)(4).
19. [ ] A second copy of the English language translation of the international application under 35 U S.C 154(d)(4).
20. [X] Other items or information:
    a. PCT Request, b. International Publication
    c. International Publication, with search report
    d. Notification of Receipt of Demand
    e. Notification of Transmittal of the Search Report
    f. Response to the International Search Report
    g. Written Opinion, h. Response to Written Opinion
    i. Form PCT/IB/304, 308, 332
    j. International Preliminary Examination Report
    k. Return Postcard

Express Mail Label No.: EL551802044 1US
Date of Deposit: FEBRUARY 7, 2002
I hereby certify that this correspondence is being deposited with the United States Postal Service "Express Mail Post Office to Addressee" service under 37 CFR §1.10 on the date indicated above and is addressed to the Assistant Commissioner for Patents, Washington, D.C. 20231.

Russ A. Griffith
Signature of person mailing paper or fee

JC13 Rec'd PCT/PTO 07 FEB 2002

U.S. APPLICATION NO (if known, see 37 CFR 1.5)	INTERNATIONAL APPLICATION NO	ATTORNEY'S DOCKET NUMBER
	PCT/US/00/22725	7024-509

21. ☒ The following fees are submitted.

BASIC NATIONAL FEE (37 CFR 1.492 (a) (1) - (5)):		CALCULATIONS PTO USE ONLY
Neither international preliminary examination fee (37 CFR 1.482) nor international search fee (37 CFR 1.445(a)(2)) paid to USPTO and International Search Report not prepared by the EPO or JPO	$1040.00	
International preliminary examination fee (37 CFR 1.482) not paid to USPTO but International Search Report prepared by the EPO or JPO	$890.00	
International preliminary examination fee (37 CFR 1.482) not paid to USPTO but international search fee (37 CFR 1.445(a)(2)) paid to USPTO	$740.00	
International preliminary examination fee (37 CFR 1.482) paid to USPTO but all claims did not satisfy provisions of PCT Article 33(1)-(4)	$710.00	
International preliminary examination fee (37 CFR 1.482) paid to USPTO and all claims satisfied provisions of PCT Article 33(1)-(4)	$100.00	
ENTER APPROPRIATE BASIC FEE AMOUNT =		$ 100--
Surcharge of **$130.00** for furnishing the oath or declaration later than ☒ 20 ☒ 30 months from the earliest claimed priority date (37 CFR 1.492(e)).		$ 130--

CLAIMS	NUMBER FILED		NUMBER EXTRA	RATE	
Total claims	83	- 20 =	63	x $18.00	$ 1134--
Independent claims	14	- 3 =	11	x $84.00	$ 924--
MULTIPLE DEPENDENT CLAIM(S) (if applicable)				+ $280.00	$ --
TOTAL OF ABOVE CALCULATIONS =					$ 2288
☒ Applicant claims small entity status. See 37 CFR 1.27. The fees indicated above are reduced by 1/2.				+	$ 1144--
SUBTOTAL =					$ 1144--
Processing fee of **$130.00** for furnishing the English translation later than ☐ 20 ☐ 30 months from the earliest claimed priority date (37 CFR 1.492(f)).					$ --
TOTAL NATIONAL FEE =					$ 1144--
Fee for recording the enclosed assignment (37 CFR 1.21(h)). The assignment must be accompanied by an appropriate cover sheet (37 CFR 3.28, 3.31). **$40.00** per property +					$ --
TOTAL FEES ENCLOSED =					$ 1144--
				Amount to be refunded:	$
				charged:	$

a. ☒ A check in the amount of $ 1144.00 to cover the above fees is enclosed.

b. ☐ Please charge my Deposit Account No. 23-3030 in the amount of $ _____ to cover the above fees. A duplicate copy of this sheet is enclosed.

c. ☒ The Commissioner is hereby authorized to charge any additional fees which may be required, or credit any overpayment to Deposit Account No 23-3030. A duplicate copy of this sheet is enclosed

d. ☐ Fees are to be charged to a credit card **WARNING:** Information on this form may become public. **Credit card information should not be included on this form.** Provide credit card information and authorization on PTO-2038.

NOTE: Where an appropriate time limit under 37 CFR 1.494 or 1.495 has not been met, a petition to revive (37 CFR 1.137 (a) or (b)) must be filed and granted to restore the application to pending status.

SEND ALL CORRESPONDENCE TO

Gregory B. COY
WOODARD, EMHARDT, NAUGHTON, MORIARTY & MCNETT
Bank One Center/Tower, Suite 3700
111 Monument Circle
Indianapolis, Indiana 46204 US

SIGNATURE

Gregory B. COY
NAME

#40,967
REGISTRATION NUMBER

METHODS AND COMPOSITIONS FOR REGULATING DEVELOPMENTAL IDENTITY

This invention was made with government support under grant number DE-FG02-97ER20133 awarded by the U.S. Department of Energy, Office of Basic Energy Sciences. The Government has certain rights in the invention.

CROSS-REFERENCE TO RELATED APPLICATIONS

The present application claims the benefit of U.S. Provisional Application Serial Number 60/149,975, filed on August 20, 1999, which is hereby incorporated by reference in its entirety.

BACKGROUND OF THE INVENTION

The present invention relates generally to methods of transforming host cells with nucleic acid encoding proteins involved in regulating developmental identity. For example, methods are also provided that include regulating embryonic identity, as well as other steps in the developmental process, especially in plants. The invention further relates to recombinant nucleic acid molecules, plant cells and transgenic plants that may be advantageously used in the methods of the present invention.

During the final stages of embryo development in angiosperms, the embryo accumulates massive amounts of nutrient storage reserves and then undergoes programmed desiccation and transition to dormancy [West, M.A. and Harada, J. J. (1993) *Plant Cell* 5: 1361-1369; Goldberg, R.B. et al. (1994) *Science* 266: 605-614; Kigel, J. and Galili, G. (Eds.) (1995) *Seed development and germination, New York,* M. Dekker ; McCarty, D. R. (1995) *Annu. Rev. Plant Phys.* 46: 71-93]. The embryo may remain dormant for extended periods of time. The quiescent embryo emerges from dormancy and undergoes post-embryonic vegetative development in response to one or more endogenous and exogenous cues that may vary

from one species to another. The regulatory processes that control the transition from the late stages of embryo development to vegetative growth and development are poorly characterized.

LEC1 appears to play a key role in regulating embryo development in *Arabidopsis* [Meinke, D. W. (1992) *Science* 258: 1647-1650; Meinke, D. W. (1994) *Plant Cell 6:* 1049-1064; West, M.A.L. et al. (1994) *Plant Cell 6:* 1731-1745; Parcy, F. et al. (1997) *Plant Cell* 9:1731-1745; Lotan, T. (1998) *Cell* 93(7):1195-205]. Seeds of *lec1* mutants exhibit numerous phenotypes, including defects in expression of maturation-specific genes, desiccation intolerance, premature germination, and abnormal expression of post-embryonic characteristics in cotyledons. *LEC1* encodes a transcription factor, the HAP3 subunit of a CCAAT box-binding factor [Lotan, T. (1998) Cell 93(7): 1195-205]. The *LEC1* transcript is expressed only in seeds, and can be detected in the embryo as early as the two-cell stage [Lotan, T. (1998) *Cell* 93(7):1195-205]. Expression of the *LEC1* gene in non-embryonic tissues is sufficient to cause expression of embryonic differentiation characteristics [Lotan, T. (1998) *Cell* 93(7):1195-205].

The ability of the growth regulator gibberellin (GA) to promote germination of seeds of numerous plant species has been demonstrated through the use of chemical inhibitors of GA biosynthesis and the characterization of mutants defective in gibberellin biosynthesis [Ritchie, S. and Gilroy, S. (1998) *New Phytol* 140:363-383]. Very little is known about the mechanism by which GA promotes germination. Genes that exhibit GA-dependent transcription are known, and the ability of GA to regulate transcription of genes in the aleurone layer of germinating cereal grains has been extensively characterized [Huttly, A. K. and Phillips, A. L. (1995) *Physiol Plant* 95:310-317; Jacobsen, J.V. et al. (1995) *The Netherlands, Kluwer Academic Publishers* 246-271; Ritchie, S. and Gilroy, S. (1998) *New Phytol* 140: 363-383]. However, a receptor for GA has not been identified. GA plays other well-characterized roles in plant growth and development in addition to its role in germination, including promotion of

elongation and regulation of the transition to flowering [Wilson, R. N. et al. (1992) *Plant Physiol* 100: 403-408; Finkelstein, R.R. and Zeevart, J. A. D. et al. (1994) *Cold Spring Harbor Laboratory:* 523-553; Hooley, R. (1994) *Plant Mol. Biol.* 26:1529-1555; Swain, S. M., Olszewski, N. E. (1996) *Plant Physiol* 112:11-17 ; Blazquez, M. A. et al. (1997) *Development* 124*: 3835-3844;* Blazquez, M. A. et al. (1998) *The Plant Cell* 10:791-800].

The ability to regulate developmental identity, such as embryonic identity, especially in plants, allows one to produce plants that have advantageous embryonic characteristics. For example, crops may be produced that include an economically significant quantity of oil. Moreover, plants that exhibit delayed flowering or reduced height may be valuable.

Although some information regarding regulation of developmental identity is known in *Arabidopsis thaliana*, identification of other proteins involved in regulation of developmental identity in lower eukaryotes could lead to identification of similar proteins in higher eukaryotes, including humans. Moreover, identification of such proteins can lead to the identification of substances that may work together with the aforementioned proteins in regulating developmental identity. There is therefore a need for nucleic acid sequences and proteins involved in regulating developmental identity. The present invention addresses this need.

SUMMARY OF THE INVENTION

A protein that functions in regulating developmental identity has been identified in the plant *Arabidopsis thaliana*. The protein is characterized by the presence of a zinc finger domain, two chromo domains, a helicase domain, and a DNA binding domain. This is the first demonstration that proteins having such features are able to regulate developmental identity, such as, for example, by terminating a previous developmental program. Accordingly, the present invention provides purified proteins having these features, including PKL (PICKLE). The invention further provides isolated nucleic acid molecules that include nucleotide sequences encoding these functional proteins. Recombinant nucleic acid molecules are also provided that include the nucleotide sequence encoding these proteins. The nucleic acid molecules may be incorporated in a host cell. Methods of transforming host cells in order to, for example, regulate developmental identity in the cells are also provided.

In a first aspect of the invention, a method of transforming a host cell is provided that includes introducing into a host cell a nucleic acid molecule encoding a protein having at least one chromo domain, a helicase domain and a DNA binding domain. The protein is advantageously expressed in an amount sufficient to regulate developmental identity. In other forms of the invention, a method may include introducing into a host cell a nucleic acid molecule encoding a protein functioning in regulating developmental identity wherein the nucleic acid molecule or the protein has the nucleotide or amino acid sequence, respectively, as described herein.

In a second aspect of the invention, a method of transforming a host cell may include introducing into a host cell an antisense DNA or RNA molecule that includes a nucleotide sequence complementary to a length of nucleotides within either a nucleic acid molecule as described herein or within a nucleic acid molecule that encodes a protein having at least one chromo domain, a helicase domain, and a DNA binding domain as

described herein. The host cell may then be cultured under conditions effective for hybridization of the antisense DNA or RNA molecule to nucleic acid of the host to regulate developmental identity. In another form of the invention, in a method of transforming a host cell, an antisense nucleic acid molecule complementary to an RNA transcript is generated by introducing into a host cell a first nucleic acid molecule having a nucleotide sequence that is complementary to a nucleotide sequence having at least about 50% identity to a length of nucleotides within the nucleotide sequence set forth in SEQ ID NO:1. After generating the antisense nucleic acid molecule, the host cell is cultured under conditions effective for hybridization of the antisense molecule to the RNA transcript of the host cell.

In a third aspect of the present invention, methods of expressing a PKL protein are provided that include introducing into a host cell the nucleotide sequences described herein and culturing under conditions effective to achieve expression of the protein.

In a fourth aspect of the present invention, recombinant nucleic acid molecules are provided that include the nucleotide sequences encoding a protein as described herein along with a foreign promoter that is operably linked to a terminal 5' end of the nucleotide sequence. Eukaryotic host cells and transgenic plants are also provided that include the introduced nucleotide sequences described herein, as are recombinant proteins. Further provided are isolated nucleotide sequences having the nucleotide sequences described herein, including those encoding the domains described herein.

It is an object of the invention to provide nucleotide sequences encoding proteins involved in regulating developmental identity, as well as the amino acid sequences encoding the proteins.

It is a further object of the invention to provide constructs, eukaryotic cells and transgenic plants that include the introduced nucleotide sequences described herein.

it js yet another object of the invention to provide methods for utilizing the nucleotide and amino acid sequences described herein, advantageously to regulate developmental identity.

These and other objects and advantages of the present invention will be apparent from the descriptions herein.

BRIEF DESCRIPTION OF THE FIGURES

FIG 1 depicts a genetic map of the region surrounding *PKL*. Markers E11M48, E11m49 pkl, E14M59, GPA-1 and nga1126 are shown below the line, whereas the distance in cM of the locus from *pkl* is indicated above the line. The extent of YAC (CIC8H12) and BAC (T3H2) clones covering the region is illustrated.

FIG 2 depicts a Southern blot performed as described in Example 1, showing polymorphisms associated with two fast neutron-derived alleles of *PKL*. *PKL* (lane 1), *pkl-7* (lane 2), and *pkl-9* (lane 3) genomic DNA were digested with Xba I and probed with the Sal I fragment indicated in FIG. 3. The numbers to the left of the figure indicate size standards.

FIG. 3 depicts a restriction map that highlights various features of the *PKL* locus as discussed in Example 1. The relative position of four open reading frames (ORFs) (P450, clpB, PKL and 2-CR) are indicated as well as the region of genomic DNA that was found not to be altered in the fast neutron-derived PKL alleles *pkl-7* and *pkl-9*. The portion of genomic DNA that was used as a probe in FIG. 2 is indicated in addition to the fragment that was used to complement the *pkl* mutant. BamHI, SalI, BstBI, and NcoI represent respective restriction endonuclease cleavage sites.

FIGS. 4A and 4B depict complementation of *pkl* phenotype in pkl plants as discussed in Example 1. Complementation of *pkl-1* seedling (FIG. 4A) and mature *pkl-1* plant (FIG. 4B) phenotype with vector carrying *PKL* is shown. For each FIG., the plant on the left is *PKL*, the plant in the middle is *pkl-1*, and the plant on the right is *pkl-1* transformed with pJ0634, as described in Example 1, which carries the *PKL* gene. The seedlings (FIG. 4A) were grown in the presence of 10^{-8} M uniconazole-P in continuous light. The mature plants (FIG. 4B) were grown under 18 hour illumination.

FIG. 5 shows a schematic diagram illustrating the location of domains of sequence homology found in PKL and other CHD proteins from *Arabidopsis* and other species as discussed in Example 2. CHD3 proteins contain PHD zinc fingers whereas CHD1 proteins do not.

FIG. 6 depicts gel analysis of a ribonuclease protection assay as discussed in Example 2. Ribonuclease protection assays were performed to determine the level of the *PKL* transcript in the root, rosette, inflorescence, and siliques of *Arabidopsis*. To demonstrate that the probe utilized was specific for *PKL*, a ribonuclease protection assay using the same probe was performed with RNA isolated from a wild-type plant and a plant carrying a deletion allele of *PKL, pkl-9* (panel on right). A probe for the cyclophilin transcript *ROC3* was used as a positive control.

FIG. 7 depicts a gel analysis of a ribonuclease protection assay, indicating that *LEC1* is expressed in pickle roots, as discussed in Example 3. Ribonuclease protection assays were used to determine the level of the *LEC1* transcript in the rosette, silique, and root of wild-type plants as well as in the pickle root of *pkl* plants. A probe for the cyclophilin transcript *ROC3* was used as a positive control.

FIG. 8 depicts a gel analysis of a ribonuclease protection assay, indicating that *LEC1* is expressed in germinating *pkl* seeds, as discussed in Example 3. Ribonuclease protection assays were used to determine the level of the *LEC1* transcripts in wild-type (WT) and *pkl* seeds at 12, 24, and 36 hours after imbibition in the absence or presence of uniconazole-P (U*). A probe for the cyclophilin transcript *ROC3* was used as a positive control.

DESCRIPTION OF THE PREFERRED EMBODIMENTS

For the purposes of promoting an understanding of the principles of the invention, reference will now be made to preferred embodiments and specific language will be used to describe the same. It will nevertheless be understood that no limitation of the scope of the invention is thereby intended, such alterations and further modifications of the invention, and such further applications of the principles of the invention as illustrated herein, being contemplated as would normally occur to one skilled in the art to which the invention relates.

A protein that functions in regulating developmental identity has been identified in the plant *Arabidopsis thaliana*. The protein is characterized by the presence of a zinc finger domain, two chromo domains, a helicase domain, and a DNA binding domain. This is the first demonstration that proteins having such features are able to regulate developmental identity, such as, for example, by terminating a previous developmental program. Accordingly, the present invention provides purified proteins having these features, including PKL. The invention further provides isolated nucleic acid molecules that include nucleotide sequences encoding these functional proteins. Recombinant nucleic acid molecules are also provided that include the nucleotide sequence encoding these proteins. The nucleic acid molecules may be incorporated in a host cell. In other aspects of the invention, methods of transforming host cells and methods of regulating developmental identity in a host cell are also provided.

In a first aspect of the invention, purified proteins are provided that include at least one chromo domain, a helicase domain, and a DNA binding domain. In preferred forms of the invention, the protein further includes a at least one zinc finger domain and preferably two chromo domains, such as found in PKL, wherein the protein functions in regulating developmental identity. As defined herein and as known in the art, developmental identity

refers generally to the identity of a tissue during, or at a stage of, development that is brought about by expression of selected genes. For example, selected genes may be expressed in a plant that gives rise to embryonic roots, and thus the developmental identity of the root is embryonic. Furthermore, selected genes may be expressed in a plant that gives rise to seedling roots, and thus the developmental identity of the root is seedling. With specific reference to PKL in pickle roots, one or more genes that gives rise to embryonic roots and one or more genes that gives rise to seedling roots are expressed simultaneously, thus the developmental identity of the root is both embryonic and seedling. The polypeptides described herein are substantially pure (i.e., the proteins are essentially free, e.g., at least about 95% free, from other proteins with which they naturally occur). In one preferred embodiment, the amino acid sequence of a PKL protein having the domains described above, originally found in *Arabidopsis thaliana*, is set forth in SEQ ID:2.

Although the invention is described with reference to *Arabidopsis thaliana* amino acid sequences, it is understood that the invention is not limited to the specific amino acid sequence set forth in SEQ ID:2. Skilled artisans will recognize that, through the process of mutation and/or evolution, polypeptides of different lengths and having differing constituents, e.g., with amino acid insertions, substitutions, deletions, and the like, may arise that are related to, or sufficiently similar to, a sequence set forth herein by virtue of amino acid sequence homology and advantageous functionality as described herein. The term "PKL protein" is used to refer generally to a protein having the features described herein and a preferred example includes a polypeptide having the amino acid sequence of SEQ ID NO:2. Also included within this definition, and in the scope of the invention, are variants of the polypeptide which function in regulating developmental identity, as described herein. Preferred proteins are recombinant proteins.

It is well known that organisms of a wide variety of species commonly express and utilize homologous proteins, which include the insertions, substitutions and/or deletions discussed above, and yet which effectively provide similar function. For example, an amino acid sequence isolated from another species may differ to a certain degree from the sequence set forth in SEQ ID NO:2, and yet have similar functionality with respect to catalytic and regulatory function. Amino acid sequences comprising such variations are included within the scope of the present invention and are considered substantially or sufficiently similar to a reference amino acid sequence. Although not being limited by theory, it is believed that the identity between amino acid sequences that is necessary to maintain proper functionality is related to maintenance of the tertiary structure of the polypeptide such that specific interactive sequences will be properly located and will have the desired activity. Although it is not intended that the present invention be limited by any theory by which it achieves its advantageous result, it is contemplated that a polypeptide including these interactive sequences in proper spatial context will have good activity, even where alterations exist in other portions thereof.

In this regard, a variant of the multi-domain protein described herein, such as a PKL protein variant, is expected to be functionally similar to that set forth in SEQ ID NO:2, for example, if it includes amino acids which are conserved among a variety of species or if it includes non-conserved amino acids which exist at a given location in another species that expresses a functional PKL protein.

Another manner in which similarity may exist between two amino acid sequences is where a given amino acid of one group (such as a non-polar amino acid, an uncharged polar amino acid, a charged polar acidic amino acid or a charged polar basic amino acid) is substituted with another amino acid from the same amino acid group. For example, it is known that the uncharged polar amino acid serine may commonly be substituted with the uncharged polar amino acid threonine in a polypeptide without

substantially altering the functionality of the polypeptide. If one is unsure whether a given substitution will affect the functionality of the enzyme, then this may be determined without undue experimentation using synthetic techniques and screening assays known in the art.

The invention therefore also encompasses amino acid sequences similar to the amino acid sequences set forth herein that have at least about 30% identity thereto and function in regulating developmental identity. Preferably, inventive amino acid sequences have at least about 50% identity, further preferably at least about 70% identity, more preferably at least about 80% identity and most preferably at least about 90% identity to these sequences.

In preferred embodiments, the invention also encompasses amino acid sequences similar to the amino acid sequences making up polypeptides having the domains described herein. For example, the invention encompasses amino acid sequences that have at least about 50%, preferably at least about 70% and more preferably at least about 90% identity to a first chromo domain from amino acid 115 to amino acid 151 or a second chromo domain extending from amino acid 191 to amino acid 227, at least about 50%, preferably at least about 70%, and more preferably at least about 90% identity to a helicase domain extending from amino acid 293 to amino acid 739, and at least about 50%, preferably at least about 70% and more preferably at least about 90% identity to a DNA binding domain extending from amino acid 1069 to amino acid 1095, and combinations thereof, all as set forth in SEQ ID NO:2. The invention further encompasses amino acid sequences, in addition to those amino acid sequences described above, that have at least about 50%, preferably at least about 70% and more preferably at least about 90% identity to the zinc finger domain amino acid sequence from amino acid 49 to amino acid 96.

Percent identity may be determined, for example, by comparing sequence information using the advanced BLAST computer program, version 2.0, available from the National Institutes of Health. The BLAST

program is based on the alignment method of Karlin and Altschul, *Proc. Natl. Acad. Sci. USA* 87:2264-68 (1990) and as discussed in Altschul, et al., *J. Mol. Biol.* 215:403-10 (1990); Karlin and Altschul, *Proc. Natl. Acad. Sci. USA* 90:5873-7 (1993); and Altschul et al. (1997) *Nucleic Acids Res.* 25:3389-3402. Briefly, the BLAST program defines identity as the number of identical aligned symbols (i.e., nucleotides or amino acids), divided by the total number of symbols in the shorter of the two sequences. The program may be used to determine percent identity over the entire length of the proteins being compared. Known default parameters are typically used, in addition to the following user-defined parameters for the BLAST program, blastp: (1) Expect value of 10.0; (2) gap penalties: Existence 11, Extension 1; and (3) scores for matched and mismatched amino acids found in the BLOSUM62 matrix as described in Henikoff, S. and Henikoff, J.G. (1992) *Proc. Natl. Acad. Sci. USA* 89:10915-10919;Pearson, W.R. (1995) *Prot. Sci.* 4:1145-1160; and Henikoff, S. and Henikoff, J.G. (1993) *Proteins* 17:49-61. The program also uses an SEG filter to mask-off segments of the query sequence as determined by the SEG program of Wootton and Federhen (1993) *Computers and Chemistry* 17:149-163.

In another aspect of the invention, isolated nucleic acid molecules, originally isolated from *Arabidopsis thaliana*, are provided that encode a functional PKL protein that functions in regulating developmental identity, especially in plants. The nucleotide sequence is set forth in SEQ ID NO:1 wherein the coding sequence is shown from nucleotide 1 to nucleotide 4152 or nucleotide 4155. It is preferred that the nucleotide sequence includes at least one of the nucleotide sequences spanning nucleotides 343 to 453 or 571 to 681, nucleotides 877 to 2217 and 3205 to 3285 in SEQ ID NO:1, which represent nucleotide sequences encoding a first chromo domain, a second chromo domain, a helicase domain and a DNA binding domain, respectively. In other forms of the invention, the nucleotide sequence further includes, in addition to the nucleotide sequences recited above, nucleotide sequences spanning nucleotides 145

to 288 in SEQ ID NO:1, which represent a nucleotide sequence encoding a zinc finger domain. It is not intended that the present invention be limited to these exemplary nucleotide sequences, but include sequences having substantial similarity thereto and sequences which encode variant forms of functional PKL protein as discussed above and as further discussed below.

The term "isolated nucleic acid," as used herein, is intended to refer to nucleic acid which is not in its native environment. For example, the nucleic acid is separated from other contaminants that naturally accompany it, such as proteins, lipids and other nucleic acid sequences. The term includes nucleic acid which has been removed or purified from its naturally-occurring environment or clone library, and further includes recombinant or cloned nucleic acid isolates and chemically synthesized nucleic acid.

The term "nucleotide sequence," as used herein, is intended to refer to a natural or synthetic sequential array of nucleotides and/or nucleosides, including deoxyribonucleic acid (DNA) and ribonucleic acid (RNA), and derivatives thereof. The terms "encoding" and "coding" refer to the process by which a nucleotide sequence, through the mechanisms of transcription and translation, provides the information to a cell from which a series of amino acids can be assembled into a specific amino acid sequence to produce a functional polypeptide, such as, for example, an active enzyme or other protein that has a specific function. The process of encoding a specific amino acid sequence may involve DNA sequences having one or more base changes (i.e., insertions, deletions, substitutions) that do not cause a change in the encoded amino acid, or which involve base changes which may alter one or more amino acids, but do not eliminate the functional properties of the polypeptide encoded by the DNA sequence.

It is therefore understood that the invention encompasses more than the specific exemplary nucleotide sequence of PKL. For example, nucleic acid sequences encoding variant amino acid sequences, as discussed above, are within the scope of the invention. Modifications to a sequence, such as deletions, insertions, or substitutions in the sequence, which

produce "silent" changes that do not substantially affect the functional properties of the resulting polypeptide molecule are expressly contemplated by the present invention. For example, it is understood that alterations in a nucleotide sequence which reflect the degeneracy of the genetic code, or which result in the production of a chemically equivalent amino acid at a given site, are contemplated. Thus, a codon for the amino acid alanine, a hydrophobic amino acid, may be substituted by a codon encoding another less hydrophobic residue, such as glycine, or a more hydrophobic residue, such as valine, leucine, or isoleucine. Similarly, changes which result in substitution of one negatively charged residue for another, such as aspartic acid for glutamic acid, or one positively charged residue for another, such as lysine for arginine, can also be expected to produce a biologically equivalent product.

Nucleotide changes which result in alteration of the N-terminal and C-terminal portions of the encoded polypeptide molecule would also not generally be expected to alter the activity of the polypeptide. In some cases, it may in fact be desirable to make mutations in the sequence in order to study the effect of alteration on the biological activity of the polypeptide. Each of the proposed modifications is well within the routine skill in the art.

In one preferred embodiment, the nucleotide sequence has substantial similarity to the sequence set forth in SEQ ID:1, especially from nucleotide 1 to nucleotide 4152 or 4155, preferably at least one of the sequences spanning nucleotides 343 to 453 or 571 to 681, nucleotides 877 to 2217 and 3205 to 3285 in SEQ ID NO:1, and variants described herein. In other forms of the invention, the nucleotide sequence, in addition to having substantial similarity to the above-recited sequences, further has substantial similarity to the nucleotide sequence spanning nucleotides 145 to 288. The term "substantial similarity" is used herein with respect to a nucleotide sequence to designate that the nucleotide sequence has a sequence sufficiently similar to a reference nucleotide sequence that it will

hybridize therewith under moderately stringent conditions. This method of determining similarity is well known in the art to which the invention pertains. Briefly, moderately stringent conditions are defined in Sambrook et al. (Eds.), *Molecular Cloning: A Laboratory Manua*l, 2nd ed. Vol. 1, pp. 101-104, Cold Spring Harbor Laboratory Press (1989) as including the use of a prewashing solution of 5X SSC (a sodium chloride/sodium citrate solution), 0.5% sodium dodecyl sulfate (SDS), 1.0 mM ethylene diaminetetraacetic acid (EDTA) (pH 8.0) and hybridization and washing conditions of 55°C, 5x SSC. A further feature of the polynucleotide is that it encodes a polypeptide having similar functionality to the PKL protein described herein, i.e., functioning to regulate developmental identity.

In yet another embodiment, nucleotide sequences having selected percent identities to the nucleotide sequence set forth in SEQ ID:1, especially with respect to the coding sequence from nucleotide 1 to nucleotide 4152 or nucleotide 4155 are provided. In one preferred form, nucleotide sequences are provided that have at least about 50% identity, preferably at least about 60% identity, more preferably at least about 80% identity, and most preferably at least about 90% identity to the nucleotide sequence set forth in SEQ ID:1, especially from nucleotide 1 to nucleotide 4152 or nucleotide 4155. In other forms of the invention, nucleotide sequences are provided that have at least about 50%, preferably at least about 60% identity, more preferably at least about 80% identity, and most preferably at least about 90% identity to a nucleotide sequence spanning nucleotides 145 to 288, at least one of the sequences spanning nucleotides 343 to 453 or 571 to 681, nucleotides 877 to 2217 and 3205 to 3285 in SEQ ID NO:1. A further feature is that the nucleotide sequence set forth in SEQ ID:1 encodes a protein that functions in regulating developmental identity.

The percent identity may be determined, for example, by comparing sequence information using the advanced BLAST computer program, version 2.0, as described above with reference to amino acid identity.

Known default parameters are typically used, in addition to the following user-defined parameters for blastn: (1) gap penalties: Existence 11, Extension 1; and (2) scores for matched and mismatched nucleotides found in the blastn matrix as described in Altschul, S.F. et al. (1997) *Nucleic Acids Res.* 25:3389-3402 and Zhang, J. (1997) *Genome Res.* 7:649-656.

A suitable DNA sequence may be obtained by cloning techniques using cDNA libraries. For example, *Arabidopsis thaliana* cDNA libraries are available commercially or may be constructed using standard methods known in the art. Suitable nucleotide sequences may be isolated from DNA libraries obtained from a wide variety of species by means of nucleic acid hybridization or polymerase chain reaction (PCR) procedures, using as probes or primers nucleotide sequences selected in accordance with the invention, such as those set forth in SEQ ID:1, nucleotide sequences having substantial similarity thereto, or portions thereof.

Alternately, a suitable sequence may be made by other techniques which are well known in the art. For example, nucleic acid sequences encoding a functional PKL protein, or variant thereof, may be constructed by recombinant DNA technology, for example, by cutting or splicing nucleic acids using restriction enzymes and DNA ligase. Furthermore, nucleic acid sequences may be constructed using chemical synthesis, such as solid-phase phosphoramidate technology. PCR may be used to increase the quantity of nucleic acid produced. Moreover, if the particular nucleic acid sequence is of a length which makes chemical synthesis of the entire length impractical, the sequence may be broken up into smaller segments which may be synthesized and ligated together to form the entire desired sequence by methods known in the art.

In another aspect of the invention, PKL polypeptides functioning in regulating developmental identity and having the amino acid sequences encoded by nucleotide sequences having substantial similarity to the nucleotide sequences described above are also provided.

In a further aspect of the invention, recombinant nucleic acid molecules, or recombinant vectors, are provided. In one embodiment, the nucleic acid molecules include a nucleotide sequence that has the selected percent identities, or substantial similarity, both as described herein, to the nucleotide sequence, or selected regions thereof, set forth in SEQ ID NO:1. In other forms of the invention, the nucleic acid molecules include a nucleotide sequence encoding a functional PKL protein. The protein produced has the amino acid sequence set forth in SEQ ID:1, or variants thereof as described above.

Recombinant vectors may be constructed by incorporating the desired nucleotide sequence within a vector according to methods well known to the skilled artisan and as described for example, in Sambrook et al. (Eds.), *Molecular Cloning: A Laboratory Manual*, 2nd Edition, Cold Springs Laboratory, Cold Spring Harbor, New York (1989). A wide variety of vectors are known that have use in the invention. For example, various plasmid and phage vectors are known that are ideally suited for use in the invention. For example, pBluescript, pGEM and pUC may be used in the invention. In preferred embodiments wherein the host cells are plants, the vector may be a T-DNA vector. Representative T-DNA vector systems are discussed in the following publications: An et al., (1986) *EMBO J.* 4:277; Herrera-Estrella et al., (1983) *EMBO J.* 2:987; Herrera-Estrella et al., (1985) in *Plant Genetic Engineering*, New York: Cambridge University Press, p. 63.

In one embodiment, the desired recombinant vector may be constructed by ligating DNA linker sequences to the 5' and 3' ends of the desired nucleotide insert, cleaving the insert with a restriction enzyme that specifically recognizes sequences present in the linker sequences and the desired vector, cleaving the vector with the same restriction enzyme, mixing the cleaved vector with the cleaved insert and using DNA ligase to incorporate the insert into the vector as known in the art.

The vectors may include other nucleotide sequences, such as those encoding selectable markers, including those for antibiotic resistance or color selection. The vectors also preferably include a promoter nucleotide sequence. The desired nucleic acid insert is preferably operably linked to the promoter. A nucleic acid is "operably linked" to a another nucleic acid sequence, such as a promoter sequence, when it is placed in a specific functional relationship with the other nucleic acid sequence. The functional relationship between a promoter and a desired nucleic acid insert typically involves the nucleic acid and the promoter sequences being contiguous such that transcription of the nucleic acid sequence will be facilitated. Two nucleic acid sequences are further said to be operably linked if the nature of the linkage between the two sequences does not (1) result in the introduction of a frame-shift-mutation; (2) interfere with the ability of the promoter region sequence to direct the transcription of the desired nucleotide sequence, or (3) interfere with the ability of the desired nucleotide sequence to be transcribed by the promoter sequence region. Typically, the promoter element is generally upstream (i.e., at the 5' end) of the nucleic acid insert coding sequence.

A wide variety of promoters are known in the art, including cell-specific promoters, inducible promoters, and constitutive promoters. The promoters may further be selected such that they require activation by activating elements known in the art, so that production of the protein encoded by the nucleic acid sequence insert may be regulated as desired. Preferred promoters are foreign promoters. A "foreign promoter" is defined herein to mean a promoter other than the native, or natural, promoter which promotes transcription of a length of DNA.

The promoters may be of viral, bacterial or eukaryotic origin, including those from plants, plant viruses and animals. As an example, the promoter may be of viral origin, including a cauliflower mosaic virus promoter (CaMV), such as CaMV 35S or 19S, a figwort mosaic virus promoter (FMV 35S), or the coat protein of tobacco mosaic virus (TMV).

Promoters of bacterial origin include the octopine synthase promoter, the nopaline synthase promoter and other promoters derived from native Ti plasmid as discussed in Herrera-Estrella et al., *Nature*, 303:209-213 (1983). Promoters of animal origin include SV40 and CMV.

The vectors may further include other regulatory elements, such as enhancer sequences, which cooperate with the promoter to achieve transcription of the nucleic acid insert coding sequence. By "enhancer" is meant nucleotide sequence elements which can stimulate promoter activity in a cell, such as a bacterial or eukaryotic host cell.

Moreover, the vectors may include another nucleotide sequence insert that encodes a protein that may aid in purification of the desired protein encoded by the desired nucleotide sequence. The additional nucleotide sequence is positioned in the vector such that a fusion, or chimeric, protein is obtained. For example, a PKL protein may be produced having at its C-terminal end linker amino acids, as known in the art, joined to the other protein. The additional nucleotide sequence may include, for example, the nucleotide sequence encoding glutathione-S-transferase (GST). After purification procedures known to the skilled artisan, the additional amino acid sequence is cleaved with an appropriate enzyme. For example, if the additional amino acid sequence is that of GST, then thrombin is used to separate the PKL protein from GST. The PKL protein may then be isolated from the other proteins, or fragments thereof, by methods known in the art.

The recombinant vectors may be used to transform a host cell. Such methods include, for example, those described in Sambrook et al. (Eds.), *Molecular Cloning: A Laboratory Manual*, 2nd Edition, Cold Springs Laboratory, Cold Spring Harbor, New York (1989). Once the desired nucleic acid has been introduced into the host cell, the host cell may produce the PKL protein, or variants thereof, as described above. Accordingly, in yet another aspect of the invention, a host cell is provided that includes the recombinant vectors described above.

A wide variety of host cells may be used in the invention, including prokaryotic and eukaryotic host cells. Bacterial host cells such as *Escherichia coli*, HB 101 and XL-1 blue may be advantageously used in the present invention. Typical eukaryotic host cells include animal host cells, such as NIH 3T3, NIH 293, COS, PCK and HeLa, and plant host cells, such as *Arabidopsis*, maize and tobacco protoplasts.

In yet another aspect of the invention, methods of producing functional PKL proteins as described above are provided. In one embodiment, the method includes providing a nucleotide sequence described above, or variants thereof, that encodes a functional PKL protein that regulates developmental identity in a host cell, and introducing the nucleotide sequence into a host cell, as described above. The desired nucleotide sequence may be advantageously incorporated into a vector to form a recombinant vector. The recombinant vector may then be introduced into a host cell according to known procedures in the art. Such host cells are then cultured under conditions, well known to the skilled artisan, effective to achieve expression of the PKL polypeptide. The PKL polypeptide may then be purified using conventional techniques.

In a further aspect of the invention, methods for transforming a host cell, which preferably allows for regulation of developmental identity, are provided. In one form of the invention, a method includes introducing into a host cell a nucleic acid molecule encoding a protein having at least one chromo domain, a helicase domain and a DNA binding domain, wherein the protein functions in regulating developmental identity. In preferred embodiments, the protein further may include at least one zinc finger domain, and further preferably includes two chromo domains. In more preferred embodiments, the protein is PKL, or a PKL variant, as described herein. The various domains may be encoded by a nucleotide sequence having selected percent identities, or substantial similarity, both as defined above, to the nucleotide sequence set forth in SEQ ID NO:1, or portions

thereof as described herein. The host cell may be cultured under conditions effective for production of said protein.

In preferred embodiments, an amount of protein is produced that is effective in regulating developmental identity. For example, the protein may regulate the transition from embryonic to post-embryonic development. In plants, for example, the protein preferably regulates the transition from an embryonic state to a seedling state. Although not being limited by theory, it is believed that PKL, or variants thereof, may act as a chromatin remodeling factor to repress transcription of LEC1, a protein that plays a role in regulating embryo development in *Arabidopsis thaliana*.

In yet other forms of the invention, the method described above may include introducing into the host cell a nucleotide sequence encoding the various domains discussed above that have at least the selected percent identities to the amino acid sequence set forth in SEQ ID NO:1 described herein.

Although the methods described herein may be performed to promote the transition from an embryonic state to a post-embryonic state, it may be advantageous in performing the methods described herein to allow the embryonic state to perpetuate after germination by altering the activity, or decreasing the production of, the protein. For example, inactivation of PKL, or variants thereof, in crops with large roots, such as radishes or turnips, may lead to production of roots that contain an economically significant amount of oil. Moreover, such inactivation may also lead to delayed flowering in plants, or to reduced height or expression of vegetative characteristics in plants, including inflorescences. In animal cells, especially mammalian cells such as human cells, altering the activity of PKL may aid in expressing particular differentiation attributes and regulation of PKL activity may have therapeutic value in human disease. As another example, regulation of PKL activity may be a convenient method to immortalize cells by inducing expression of stem cell differentiation characteristics. Alternatively, PKL genes may be potential

oncogenes, and loss of their function may lead to cells inappropriately expressing stem cell characteristics. Similarly, some teratomas may be caused by inactivation of PKL genes, causing the inappropriate expression of various differentiation programs.

Accordingly, in other forms of the invention, proteins are provided having the features described herein that are modified so that the embryonic state may be maintained after entry into the post-embryonic state. In one form of the invention applied to, for example, plant host cells, a method of regulating developmental identity may include *in vivo* mutagenesis of the gene present in the host genome that encodes the protein described herein in order to alter its activity to provide the desired results. For example, a plant may be mutated by methods known to the skilled artisan, including chemical methods and homologous recombination methods. Moreover, other methods include use of interference RNA, T-RNA and fast-neutron mutagenesis. All of these methods are well known to the art, and may be found, for example, in Koncz et al. (Eds.) *Methods in Arabidopsis Research*, World Scientific Publishing Co. (1992).

In yet other forms of the invention, one of the domains, or other regions of the proteins described herein, may be deleted in order to inactivate, or otherwise decrease the activity of, the PKL protein produced. It is realized that all, or a portion of one or more domains may be deleted by methods that include PCR mutagenesis and recombinant DNA technology, as known in the art and as exemplified in, for example, Sambrook et al. (Eds.), *Molecular Cloning: A Laboratory Manua*l, 2nd ed., Cold Spring Harbor Laboratory Press (1989).

In yet other forms of the invention, a method of transforming a host cell, preferably to regulate developmental identity, includes introducing into a host cell an antisense nucleotide sequence having a nucleotide sequence complementary to a length of nucleotides within a nucleic acid molecule as described herein. For example, the nucleic acid molecule may encode a protein having at least one chromo domain, a helicase domain, and a DNA

binding domain, or other protein as described herein, such as one having an amino acid sequence having the selected percent identities to the various domains in SEQ ID NO:2 as described herein, including the zinc finger domain. The cell is typically cultured for a time period and under conditions effective for hybridization of the antisense nucleic acid sequence to nucleic acid of the host. The antisense nucleic acid sequence may be DNA or RNA. The length of nucleotides the antisense nucleotide sequence may be complementary to is typically a length sufficient for hybridization to the target nucleic acid sequence so that transcription and/or translation will be substantially inhibited and/or production of a functional protein will be substantially stopped or otherwise substantially decreased. For example, antisense nucleotide sequence may be at least about 25 nucleotides long, and may further be about 50 to about 4200 nucleotides long, preferably about 100 to about 1000 nucleotides long, and further more preferably about 200 to about 500 nucleotides long. In preferred forms of the invention, the antisense nucleic acid sequence may be complementary to, for example, a region from about nucleotide 2 to about nucleotide 331 set forth in SEQ ID NO:1. In other preferred forms of the invention, the antisense nucleic acid sequence may be complementary to a region from about nucleotide 3330 to about nucleotide 3710 in SEQ ID NO:1.

In yet another form of a method of transforming a host cell, a method may include introducing into the host cell a vector that includes a nucleic acid molecule that may be used to generate a nucleic acid molecule, such as an antisense RNA molecule, that will bind to the endogenous transcript in order to inhibit translation of the transcript and to target the transcript for degradation. In one form, the method may include introducing into the host cell a vector that includes length of nucleotides within the nucleotide sequence shown in SEQ ID NO:1 along with the same nucleotides in an antisense orientation. As an example, the host cell may be transformed with a construct that includes, in the following order, a promoter, operably linked to, for example, a PKL fragment as described herein in the sense

orientation, an intron, the same PKL fragment in an antisense orientation and a terminator, as described in Example 5. A double-stranded transcript may be generated in the host cell after the intron is spliced out, which may then generate a complementary RNA molecule through a double-stranded RNA-dependent RNA polymerase. This complementary RNA molecule may then bind to the endogenous transcript, such as the messenger RNA (mRNA), and target it for degradation, as known in the art.

Reference will now be made to specific examples illustrating the molecules, cells and methods above. It is to be understood that the examples are provided to illustrate preferred embodiments and that no limitation to the scope of the invention is intended thereby.

EXAMPLE 1

Cloning of PKL

Plant Material and Media

The *pkl-1* mutation was isolated from an EMS-mutagenized population of the Col ecotype [Ogas, J. et al. (1997) *Science* 277: 91-94]. The *pkl-7, pkl-8,* and *pkl*-9 alleles were isolated from a fast neutron-mutagenized population of the Col ecotype that was obtained from Lehle Seeds (http://www.arabidopsis.com/ cat. # M2F-01A-04). Plants were grown as described previously [Ogas, J. et al. (1997) Science 277: 91-94].

Cloning of PKL

pkl-1 plants of the Col ecotype were crossed to plants of the Landsberg *erecta* type to generate a mapping population, and 300 F2 progeny expressing the pickle root phenotype were isolated. DNA was isolated from these progeny using a protocol described by [Liscum, E. and Oeller, P. W. (1999) *Genome Analysis, P. Offner (*Ed.), CRC Press, Boca Raton, FL, in press]. The SSLP markers used are described at http://genome.bio.upenn.edu/ SSLP_info/ SSLP.html, and the PCR analysis of the markers was done as previously described [Bell, C. J. and

Ecker, J. R. (1994) *Genomics* 19:137-144]. The AFLP analysis was performed as described by Liscum, E. and Oeller, P. W. (1999) *Genome Analysis, P. Offner* (Ed.),. CRC Press, Boca Raton, FL, in press]. The AFLP primers used for mapping analysis were as follows: the basic EcoRI primer is 5'-AGA CTG CGT ACC ATT TCx y-3' (where x and y indicate base pairs added for specificity), shown in SEQ ID NO:3, and the basic MseI primer is 5'-GAT GAG TCC TGA GTA Axy z-3' (where x, y, and z indicate base pairs added for specificity), shown in SEQ ID NO:4. E11M48 denotes the primer pair EcoRI-AA and MseICAC, E11M49 denotes the primer pair EcoRI-AA and MseI-CAG, and EI4M59 denotes the primer pair EcoRI-AT and MseICTA [Alonso-Blanco, C. et al. (1998) *Plant J.* 14: 259-271].

To identify polymorphisms in the fast neutron-derived alleles of *PKL,* Southern blots were performed using genomic DNA from plants and digoxigenin-labeled probes that were generated from YAC DNA using AFLP technology. DNA from YAC CIC8H12 (YAC CIC8H12 was obtained from the Arabidopsis Biological Resource Center, Columbus, Ohio) and was prepared as described [Gibson, S. I. and Somerville, C. (1992) *World Scientific:* 119-143]. Approximately 50 ng of CIC8H12 DNA was utilized in a restriction and ligation reaction as described at http://carnegiedpb.stanford.edu/methods/aflp.htmi, with the following differences: the DNA was only digested with MseI, and only the MseI adaptor was ligated on. Five µl of this restriction and ligation (RAL) mixture was then used in a 100 µl digoxigenin-labeling PCR reaction (Roche Biochemicals, cat. # 1 636 090) with 100 pmol each of 6 MseI-xy primers (where x and y indicate base pairs added for specificity). The entire PCR reaction was then used to probe a Southern blot as described in the Dig User's Guide (Roche Biochemicals, cat. # 1 438 425). Random combinations of 6 MseI-xy primers were used to screen for polymorphisms in the fast neutron-derived alleles. Polymorphisms were revealed when the following 6 primers were utilized: xy = CT, GG, GC, AG, TG, AT.

To identity a bacterial artificial chromosome (BAC) that spanned the PKL locus, Southerns were performed using BAC filters and a probe generated from the AFLP marker E11M49. BAC filters representing the Arabidopsis genome were obtained from the Arabidopsis Biological Resource Center at Ohio State University (stock # CD4-25F). Southern blots were performed as described in the Dig User's Guide (Roche Biochemicals, cat. # 1 438 425). BAC T3H2 was identified as a positive, and DNA was isolated using a midiprep kit and protocol from Qiagen (cat. # 12143). Approximately 5 ng of T3H2 DNA was utilized to generate a DIG-labeled AFLP probe as described above for CIC8H12. The same 6 primers that identified polymorphisms with CIC8H12 also gave polymorphisms with T3H2. Bands that were polymorphic in fast neutron-derived alleles of PKL were then subcloned from T3H2 and the DNA sequence was determined using an ABI 310.

Complementation of pkl mutant

A BstBI - NcoI 11.9 kb genomic fragment that spanned the predicted *CHD* gene was subcloned into the plant transformation vector pCambia 3300 (CSIRO, Canberra) using the BstXI and XbaI sites to generate pJ0634. pkl-1 and pkl-7 plants were transformed with both empty vector and pJ0634 using an *in planta* transformation protocol with the *Agrobacterium tumefaciens* strain GV3101 [Bechtold, N. et al. (1993) *C. R. Acad. Sci. Paris* 316: 1194-1199]. Basta was used to select for transformants of T1 progeny. Only *pkl* plants transformed with pJ0634 were complemented for the *pkl* phenotype. The T2 progeny of two independent transformants that exhibited a complemented phenotype were examined for cosegregation of basta-resistance and complementation. For both lines, basta-resistance and complementation cosegregated indicating that complementation was due to introduction of the wild-type *PKL gene*.

Results

Fast neutron-derived alleles of *PKL* were identified to facilitate the cloning of *PKL* by map-based methods. Fast neutron mutagenesis generates mutations that consist of chromosomal deletions at a high frequency [Bruggemann, E. et al. (1996) *Plant J.* 10: 755-760]. Approximately 50,000 fast neutron-mutagenized M_2 seed were screened for the pickle root phenotype in the presence of 10^{-8} M uniconazole-P, a GA biosynthetic inhibitor [Izumi, K. et al. (1985) *Plant Cell Physiol.* 26: 821-827] that increases penetrance of the pickle root phenotype [Ogas, J. et al. (1997) *Science* 277:91-94]. Three independent *pkl* mutants were identified and utilized as described below.

The *pkl* mutation was genetically mapped relative to previously mapped polymorphisms between the Col and Ler ecotypes of *Arabidopsis*. Plants carrying the *pkl-1* allele in the Col ecotype were crossed to wild-type Ler plants and 300 F_2 progeny expressing the pickle root phenotype were isolated. DNA from the 300 *pkl* F_2 was used to localize the *pkl-1* mutation by interval mapping using SSLP markers [Bell, C. J. and Ecker, J. R. (1994) *Genomics* 19: 137-144]. The *pkl* mutation mapped to chromosome 2 near the nga1126 marker (FIG. 1). Based on the analysis of 231 F_2 progeny, the *pkl-1* mutation mapped to within 1.1 cM of the SSLP marker GPA-I which had been anchored on the physical map of chromosome 2 [Wang et al. (1997) *Plant J.* 12:711-730]. Further analysis of the 231 F_2 progeny revealed that the AFLP markers [Prabhu, R. R. and Gresshoff, P. M. (1994) *Plant Mol. Biol.* 26: 105-116; Alonso-Blanco, C. et al., (1998) *Plant J.* 14: 259-271] E11M48 and E14M59 flanked *pkl-1* and were tightly linked (FIG. 1).

Based on the position of *pkl* on the physical map of chromosome 2 [Wang, M. L. et al. (1997) *Plant J.* 12: 711-730], YAC CIC8H12 was selected for further analysis. PCR analysis revealed that CIC8H12 contained the flanking markers E11M49 and E14M59 (FIG. 1), indicating that CIC8H12 spanned the *PKL* locus (data not shown). Five pools of random probes were generated from CIC8H12 by a PCR-based method.

These random probe mixtures were then used to probe Southern blots of genomic DNA isolated from wildtype plants and the three *pkl* lines generated by fast neutron mutagenesis. One of the probes revealed polymorphic bands associated with 2 of the 3 fast neutron alleles (data not shown).

We also screened for a BAC clone that spanned the *pkl* locus. The AFLP marker E11M49, which mapped 0.23 cM from *pkl,* was cloned and then used to probe BAC filters covering the Arabidopsis genome [Woo, S. - S. et al. (1994) *Nucleic Acids Res.* 22: 4922-4931; Choi, S. D. et al. (1995) *Weeds World* 2: 17-20]. Several BACs that hybridized to the clone were identified. Restriction analysis of these BACs revealed that BAC T3H2 was likely to span the *pkl* locus. T3H2 contained restriction fragments that were identical in size to the restriction fragments from wild type that were polymorphic in the fast neutron lines. A random probe mixture was generated from T3H2 by PCR utilizing the same pool of primers used to generate a random probe mixture from YAC CIC8H12. This probe mixture from T3H2 identified the same polymorphic bands in the fast neutron lines as the probe mixture from CIC8H12 (data not shown).

The nature of the lesions in the fast neutron lines was characterized in greater detail using specific probes generated from T3H2. Various DNA fragments from T3H2 were subcloned and used as probes on Southern blots of Arabidopsis genomic DNA. One of these Southerns is shown in FIG 2. The probe used for this blot was a 10 kb Sal I fragment indicated in FIG 3. Lanes 1, 2, and 3 contain genomic DNA digested with Xba I that was isolated from wild-type plants, fast neutron allele *pkl-7,* and fast neutron allele *pkl-9*, respectively. Polymorphic bands are seen in *pkl-7* and *pkl-9.* Based on Southerns such as the one shown in FIG. 2, the extent of the alterations in the genomic DNA in *pkl-7* and *pkl-9* was deduced to be as shown in FIG 3. The mutation in *pkl-7* is caused by either a translocation or an insertion whereas the mutation in *pkl-9* is caused by a large deletion.

Sequencing of the wild-type genomic DNA surrounding the *pkl-7* polymorphism indicated that only one gene is disrupted in both the *pkl-7* and *pkl-9* mutants. A 3.0 kb BamHI fragment of genomic DNA that was polymorphic in the *pkl-7* line was sequenced and a portion of a potential gene encoding a putative CHD protein was identified. Since CHD proteins can be greater than 2000 amino acids in length, 17 kb of genomic DNA was sequenced to ensure that the entire potential CHD gene was sequenced. The Genbank database was searched with the sequenced 17 kb region using the program BLASTX [Altschul, S. F. et al. (1997) *Nuc. Acids Res.* 25: 3389-3402], which translates the DNA of interest in all 6 reading frames and compares the translations to the protein database. Based on this database search, the sequenced 17 kb region contains all or part of 4 genes, as indicated in FIG. 3. These 4 genes have sequence similarity to a cytochrome P450 monooxygenase, a clpB protease, a CHD family member, and a 2-component regulator [Ogas, J. et al. (1997) *Science* 277: 91-94]. Only the gene coding for the CHD family member is disrupted in both the pkl-7 and pkl-9 mutants (FIG. 3).

Complementation analyses confirmed that *PKL* has been cloned. A binary vector, pJ0634, carrying an 11.9 kb BstBI - NcoI genomic fragment that spans the predicted *CHD* gene (FIG. 3) was constructed and transformed into *pkl* plants. pkl plants transformed with pJ0634 are complemented for all *pkl*-related phenotypes (FIG. 4), whereas *pkl* plants transformed with the vector alone are not (data not shown). Segregation analyses was done on two independent lines transformed with pJ0634 to confirm that the ability to suppress the *pkl* mutant phenotype cosegregated with the transgene (data not shown).

EXAMPLE 2

Characterization of PKL

Ribonuclease protection assays.

Ribonuclease protection assays were performed using the RPA III kit from Ambion (cat. # 1414). To generate a PKL-specific probe, a DNA fragment was generated via RT-PCR using the primers JOpr244 (5'-TGT TGA GCC AGT TAT TCA CGA-3'), (nucleotides 1725-1745 in SEQ ID NO:1) shown in SEQ ID NO:5, and JOpr247 (5'-ACC TTT CCA TCA ATT CGC TCG-3') (sequence complementary to nucleotides 1934-1914 in SEQ ID NO:1) shown in SEQ ID NO:6, and subcloned using the pGEM-T vector system (Promega, cat. # A3600) in an orientation such that the T7 promoter would produce an anti-sense transcript. This plasmid was called pJ0657. To generate a *LEC1*-specific probe, a DNA fragment was generated via PCR using the primers JOpr273 (5'CCGCTCGAGAACCCCAATGACCAGCTCAGT-3'), shown in SEQ ID NO:7 (the first 3 nucleotides are used as spacers so the restriction enzyme will cut properly, the next 6 nucleotides represent the XhoI recognition sequence and the last 21 nucleotides are nucleotides 33-53 of LEC1 cDNA sequence, Genbank Accession No. AF036684), and JOpr262 (5'-CCTTCTTCACTTATACTGACC-3'), shown in SEQ ID NO:8 (sequence complementary to nucleotides 672-652 of LEC1 cDNA sequence, Genbank Accession No. AF036684), digested with XhoI and KpnI and subcloned into pBluescript SK cut with XhoI and KpnI to produce pJ0660. To generate a ROC3-specific probe, a DNA fragment was generated via PCR using the primers JOpr276 (5'-AAGTCTACTTCGACATGACCG-3'), shown in SEQ ID NO:9 (nucleotides 65-85 of ROC3 cDNA sequence, Genbank Accession No. U40399), and JOpr277 (5'-CTTCCAGAGTCAGATCCAACC-3'), shown in SEQ ID NO:10 (sequence complementary to nucleotides 524-504 of ROC3 cDNA sequence, Genbank Accession No. U40399), and subcloned using the pGEM-T vector system in an orientation such that the T7 promoter would produce an anti-sense transcript. This plasmid was called

pJ0662. To generate ^{32}P-labeled RNA probes for RPA analysis, the T7 Maxiscript kit was used (Ambion cat. # 1312) with pJ0657, pJ0660, and pJ0662 digested with NotI. The full-length transcripts were gel-purified to reduce background. For each ribonuclease protection assay, approximately $2x10^4$ CPM of probe was added to 10 µg of total RNA [Verwoerd, T. C. et al. (1989) *Nuc. Acids Res.* 17: 2362-2362].

Results - Sequence Comparison

RT-PCR was used to clone cDNA fragments representing the entire predicted *PKL* ORF. Subsequently, a BAC that spanned the *PKL* locus, F1 3D4 (Acc# AL031369), was sequenced by another group as part of the ongoing effort to sequence the Arabidopsis genome. The sequences were identical, with the exception that some of the splice sites that were utilized to generate the *PKL* transcript were different from those predicted by the computer algorithm (the *PKL* cDNA sequence is deposited in Genbank, accession #AF185577). Analysis of the *PKL* ORF revealed that *PKL* codes for a predicted CHD3 homolog that is 1385 amino acids in length. A search of the Genbank database revealed that genomic sequence for another Arabidopsis CHD3 homolog that is located on chromosome V (Accession # AAC79140) has also been obtained by the genome project. Also, an Arabidopsis CHD1 homolog is located on chromosome IV (Accession # CAB40760). We refer to this other CHD3 homolog as *PICKLE RELATED 1 (PKR1)* and the CHD1 homolog as *PICKLE RELATED 2 (PKR2)*

PKL, PKRI, and *PKR2* contain all of the sequence domains expected of CHD proteins [Delmas, V. et al. (1993) *Proc. Natl. Acad. Sci. USA* 90: 2414-2418; Woodage, T. et al. (1997) *Proc. Natl. Acad. Sci. USA* 94: 11472-11477]. CHD proteins are defined by three domains of sequence similarity: a chromo (chromatin organization modifier) domain, a SNF2-related helicase/ATPase domain, and a DNA-binding domain. CHD3 proteins are distinguished from CHD1 proteins by the presence of another domain, a PHD zinc finger [Woodage, T. et al. (1997) *Proc. Natl. Acad.*

Sci. USA 94: 11472-11477]. FIG. 5 is a schematic of the various domains found in PKL, PKR1, PKR2 and related CHD proteins. Table 1 lists the percent similarity between domains in *PKL* and related domains in the other proteins.

Table 1. Comparison of sequence identity of various domains found in other CHD proteins to domains found in PKL. Percent identity is indicated for PHD zinc fingers (PHD), chromo domains (chromo), SNF2-related helicase/ATPase domain (helicase), and DNA binding domain (DNA). For the PHD zinc fingers, both of the PHD zinc fingers from the other CHD3 proteins are compared to the single PHD zinc finger from PKL.

	PHD#1	PHD#2	Chromo#1	Chromo#2	Helicase	DNA
Human CHD3	33	35	32	37	58	40.7
Drosophila CHD3	31.2	38	35	45	55	44
PRK1		33	24	42	51	30
PKR2			68	78	75	74
Yeast CDH1			19	35	49	49
Mouse CHD1			32	27	50	33
Length (amino acids)	48	48	37	37-38	452-469	27

Only one PHD zinc finger is found in *PKL* and *PKRI*, whereas 2 PHD zinc fingers are typically found in CHD3 proteins from other species. Based on the domains of homology identified, we have classified *PKL* and *PKR1* as CHD3 family members and *PKR2* as a CHDI family member. *PKRI* is distinguished from the other CHD3 proteins by the fact that the PHD zinc finger is located more towards the N-terminus of the protein than the PHD zinc fingers of the other CHD3 proteins.

PKL appears more similar to the putative CHD1 protein PKR2 than the putative CHD3 protein PKR1. Several pairwise comparison programs

were unable to correctly align all of the various domains of PKL and PKR1, whereas PKL and PKR2 were correctly aligned and exhibit 54% sequence identity over the entire protein. In this regard, it is interesting to note that the spacing between the SNF2-related helicase domain and the putative DNA-binding domain that is observed in PKL is more similar to that of CHD1 proteins than that of CHD3 proteins (FIG. 5).

Results-Expression of PKL

To determine where *PKL* is normally expressed, *PKL* transcript levels were analyzed. The *PKL* transcript was not detected by Northern analysis of poly(A+) mRNA of rosette leaves. This may be due to technical difficulties associated with preparation of long transcripts from plant tissues [Roesler, K. R. et al. (1994) *Plant Physiol* 105: 611-617]. Therefore, ribonuclease protection assays were used to quantitate *PKL* mRNA (FIG. 6). At this level of resolution, the *PKL* transcript was present at approximately equal levels in all tissues examined: roots (lane 1), shoots (lane 2), inflorescences (lane 3), and siliques (lane 4). This ubiquitous expression pattern is consistent with the pleiotropic shoot and root phenotypes exhibited by *pkl* plants. The *PKL* transcript was not detected when the ribonuclease protection assay was performed on RNA isolated from a plant carrying a deletion allele of *PKL, pkl-9* (lanes 5 and 6).

EXAMPLE 3

Expression of LEC1

Pickle roots are primary roots of adult plants that express embryonic differentiation traits such as expression of storage protein genes and accumulation of storage lipids [Ogas, J. et al. (1997) *Science* 277: 91-94]. These and other embryo-specific traits are thought to be under control of the *LEC1* gene, which has been proposed to be a critical regulator of

embryonic identity [Meinke, D. W. (1992) *Science* 258: 1647-1650; Meinke, D. W. et al. (1994) *Plant Cell 6*: 1049-1064; West, M. A. L. et al. (1994) *Plant Cell 6*: 1731-1745; Parcy, F. et al. (1997) *Plant Cell 9*: 1265-1277; Lotan, T. et al. (1998) *Cell* 93(7): 1195-1205]. Therefore, the possibility that the *LEC1* transcript, which is normally only expressed in seeds, was expressed in pickle roots was investigated. Ribonuclease protection assays were performed using total RNA isolated from wild-type roots and pickle roots with a *LEC1* probe and a cyclophilin probe as a control (FIG. 7). As expected, the *LEC1* transcript was detected in siliques (lane 2) but not in rosette leaves (lane 1). Although the *LEC1* transcript was not detected in wild-type roots (lane 3), expression of *LEC1* was clearly detected in pickle roots (lane 4).

Since expression of *LEC1 is* sufficient to induce expression of embryonic differentiation traits in seedlings [Lotan, T. (1998) *Cell* 93(7): 1195-1205], the presence of the *LEC1* transcript in pickle roots suggested that *LEC1* may play a key role in promoting expression of the pickle root phenotype. Penetrance of the pickle root phenotype in *pkl* seedlings is induced by treatment of seed with uniconazole-P prior to germination. If the level of the *LEC1* transcript is the limiting factor in determining the penetrance of the pickle root phenotype, then the *LEC1* transcript would be predicted to exhibit uniconazole-P dependent expression in imbibed *pkl* seeds.

It was found that the *LEC1* transcript was present in imbibed *pkl* seeds prior to germination (FIG. 8). Ribonuclease protection assays were performed using total RNA isolated from wild-type seed (lanes 1-6) and *pki* seed (lanes 7-12) with a *LEC1* probe and a cyclophilin probe as a control. Seeds were imbibed in the absence or presence of uniconazole-P for 12, 24 or 36 hours. The *LEC1* transcript is clearly present in *pkl* seeds at 24 hours and 36 hours. However, the level of the *LEC1* transcript was not elevated in *pkl* seed treated with uniconazole-P.

Analysis

PKL is a CHD3 gene

In wild-type *Arabidopsis*, many of the developmental pathways that contribute to embryo formation are not expressed in adult tissues. In *pkl* mutants, at least some aspects of this stage-specific control are lost; embryonic developmental programs such as expression of seed storage protein genes and genes involved in storage lipid deposition are expressed after germination [Ogas, J. et al. (1997) *Science* 277: 91-94]. Moreover, vegetative tissues have an abnormal capacity to spontaneously produce somatic embryos. In *pkl* seedlings, all organs generated during embryogenesis are capable of expressing embryonic identity after germination [Ogas, J. et al. (1997) *Science* 277: 91-94] (manuscript in preparation). By contrast, organs that arise post-embryonically, such as secondary roots, never express embryonic traits [Ogas, J. et al. (1997) *Science* 277: 91-94] (unpublished observations). Thus, *PKL* is apparently necessary to repress embryonic identity and contributes to the transition from embryonic to post-embryonic development.

The identification of *PKL* as a gene encoding a CHD3 protein suggests that *PKL* mediates its effects on developmental identity through regulation of chromatin architecture. CHD genes have been identified in numerous eukaryotes, and the corresponding proteins are proposed to function as chromatin remodeling factors. The name "CHD" is derived from the three domains of sequence homology found in CHD proteins [Delmas, V. (1993) *USA* 90: 2414-2418; Woodage, T. et al. (1997) *USA 94*: 11472-11477] a chromo (chromatin organization modifier) domain, a SNF2-related helicase/ATPase domain, and a DNA-binding domain. Chromo domains are proposed to function as protein-protein interaction domains [Cowell, I. G. and Austin C. A. (1997) *Biochim. Biophys. Acta* 1337:198-206] and are found in numerous chromatin-associated proteins [Koonin, E. et al. (1995) *Nuc. Acids Res.* 23: 4229-4233]. The SNF2-related helicase/ATPase domain is found in numerous proteins that exhibit different activities

towards DNA [Eisen, J. A. et al. (1995) *Nuc. Acids Res.* 23: 2715-2723]. The SNF2-related helicase/ATPase domain found in CHD genes exhibits highest sequence similarity to the SWI/SNF class of transcriptional activators, which are proposed to remodel chromatin [Hirschhorn, J. N. et al. (1992) *Genes & Dev.* 6: 2288-2298; Prelich, G. and Winston, F. et al. (1993) *Genetics* 135: 665-676; Imbalzano, A. N. et al. (1994) *Nature* 370(6489): 481-5; Kwon, H. et al. (1994) *Nature* 370(6489): 477-81; Kruger, W. et al. (1995) *Genes & Dev.* 9: 2770-2779; Owen-Hughes, T. et al. (1996) *Science* 273(5274): 513-6; Logie , C. and Peterson, C. L. (1997) *Embo J* 16(22): 6772-82] by an as yet undetermined mechanism. The DNA binding domain of the CHD proteins is most similar to that of the telobox subset of Myb-related DNA-binding motifs [Woodage, T. et al. (1997) *Proc. Natl. Acad. Sci. USA* 94: 11472-11477]. Thus, CHD proteins are a unique juxtaposition of three domains with chromatin-related activities in a single polypeptide.

At present, four CHD genes have been sequenced from *Arabidopsis*: *PKL, PKRI, PKR2* and *PKR3.* CHD proteins are separated into two classes, CHD1 and CHD3, based on domains of homology found in the proteins [Woodage, T. et al. (1997) *Proc. Natl. Acad. Sci. USA* 94: 11472-11477]. CHD3-related proteins are distinguished from CHD1-related proteins by the presence of an additional domain of homology, the PHD zinc finger [Woodage, T. et al. (1997) *Proc. Natl. Acad. Sci. USA* 94: 11472-11477]. *PKL* and *PKR1* both have a single PHD zinc finger. Based on the presence of that motif, we have classified them as CHD3 proteins. This classification brings with it certain experimental predictions; CHD3 proteins have been shown to be associated with histone deacetylases (see below). *PKR2 and PKR3* do not have a PHD zinc finger and so we have classified them as CHD1 proteins.

CHD3 proteins are thought to be involved in repression of transcription. CHD3 proteins from *Xenopus* and human have been show to be a component of a complex that contains histone deacetylase as a

subunit [Tong, J. K. et al. (1998) *Nature* 395: 917-921; Wade, P. A. et al. (1998) *Curr. Biol.* 8: 843-846; Zhang, Y. et al. (1998) *Cell* 95(2): 279-289]. Deacetylation of histones is correlated with transcriptional inactivation [Turner, B. M. (1991) *J. Cell Sci.* 99:13-20; Grunstein, M. (1997) *Nature* 389:349-352; Struhl, K. (1998) *Genes & Dev.* 12:599-606]. Thus, by virtue of CHD3 proteins being a component of a histone deacetylase complex, they would be predicted to function as repressors of transcription. In a mutant of *Drosophila* that lacks the CHD3-related gene d*Mi-2*, this prediction is borne out; homeotic genes that are normally repressed are derepressed in a d*Mi-2* mutant [Kehle, J. et al. (1998) *Science* 282(5395): 1897-1900].

There is little published evidence of the function of CHD1 proteins. Deletion of the only CHD gene in yeast, *a CHDI* gene, does not result in *a* phenotype under standard growth conditions However, *chdl* yeast exhibit increased resistance to the pyrimidine analog 6-azauracil, a phenotype which is consistent with a role for CHD1 in repression of transcription [Woodage, T. et al. (1997) *Proc. Natl Aad. Sci. USA* 94: 11472-11477].

Based on the data presented here and previously, it is proposed herein that *PKL* also functions as a repressor of transcription. In *pkl* mutants, embryo-specific genes are expressed inappropriately after germination [Ogas, J. et al. (1997) *Science* 277: 91-94]. Such derepression could be due to loss of a shared repressor of embryo-specific genes or due to inappropriate expression of a general activator of the embryo-specific genes. *LEC1* codes for a seed-specific transcription factor and is a critical activator of the embryonic developmental program [Lotan, T. (1998) *Cell* 93(7): 1195-1205]. We have shown that *LEC1* is expressed in *pkl* tissue expressing embryonic differentiation characteristics after germination.

Since expression of *LEC1* after germination is sufficient to cause expression of embryonic differentiation characteristics [Lotan, T. (1998) *Cell* 93(7): 1195-1205], one possible model to explain expression of

embryonic identity after germination in *pkl* seedlings is that *PKL* is necessary for repression of *LEC1*. We found that *LEC1* is expressed in *pkl* seeds prior to germination (FIG. 8), but the level of the *LEC1* transcript is not increased in the presence of uniconazole-P. Based on what is known about penetrance of the pickle root phenotype, *PKL*, and *LEC1*, this result is consistent with a direct role for *PKL* in repression of *LEC1* and with a substantive role for *LEC1* in generation of the pickle root phenotype. However, this result is not consistent with a role for *LEC1* as a rate-determining factor governing penetrance of the pickle root phenotype. In fact, the result strongly suggests that there is a separate factor that promotes expression of embryonic genes that is in some way repressed by GA.

PKL is a component of a GA-dependent developmental switch

Based on the characterization of the phenotype of the *pkl* plant described in this study and on the identification of *PKL* as a CHD3 gene, the following model is proposed herein to explain the role of *PKL* in regulating developmental identity during germination. Briefly, in response to a GA-dependent signal, *PKL* remodels the chromatin upstream of one or more genes that promote embryonic identity into a transcriptionally incompetent state. As a consequence of this transcriptional inactivation, expression of the embryonic developmental program is repressed after germination. In conjunction with previous observations concerning GA, the results in this study imply that GA plays two roles in germinating seeds of *Arabidopsis*. One well-established role is that GA triggers metabolic activity and activates postembryonic developmental processes. In addition, the results in this study indicate that GA plays a role in repression of embryonic developmental processes. Thus, it is proposed herein that GA acts as both a differentiation factor (promotion of the postembryonic state) and a determination factor (repression of the embryonic state) during germination. This result is surprising, especially in light of previous results

with double mutants of *Arabidopsis* defective in both ABA and GA biosynthesis [Koomeef, M. (1982) *Theor. Appl. Genet.* 61: 385-393]. Such mutants germinate in the absence of GA and do not inappropriately express embryonic differentiation characteristics after germination. One possible explanation for this apparent contradiction is that factors in addition to GA may be able to promote repression of embryo-specific genes.

It is proposed in this study that *PKL* activity is, in some way, GA-dependent. What has been observed in this study is that pickle root penetrance is GA-dependent in the absence of *PKL*. What this observation implies is that *PKL* and a factor whose activity is GA-dependent are necessary for repression of embryonic genes. The supposition that the activity of *PKL* itself is in some way GA-dependent is based on observations that the shoot phenotype of *pkl* plants is consistent with a defect in a GA signal transduction pathway [Ogas, J. et al. (1997) Science 277: 91-94]; (manuscript in preparation). Based on the conclusion that *PKL* is functioning in a GA signal transduction pathway during shoot development, it is proposed herein that the activity of *PKL* is similarly regulated by GA during germination.

The hypothesis that *PKL* remodels chromatin into a transcriptionally incompetent state is consistent with published data regarding CHD3 proteins and with the *pkl* mutant phenotype. CHD3 proteins have been shown to associate with histone deacetylase [Tong, J. K. et al. (1998) *Nature* 395: 917-921; Wade, P. A. et al. (1998) *Curr. Biol.* 8: 843-846; Zhang, Y. et al. (1998) *Cell* 95(2): 279-289], and deacetylation of histones is correlated with reduced transcription [Turner, B. M. (1991) *J. Cell Sci.* 99: 13-20; Grunstein, M. (1997) *Nature* 389: 349-352; Struhl, K. (1998) *Genes & Dev.* 12: 599-606]. In a *Drosophila* mutant lacking the CHD3-related gene M-2, homeotic genes are derepressed [Kehle, J. et al. (1998) *Science* 282(5395): 1897-1900]. In the *pkl* mutant, embryonic genes are derepressed after germination [Ogas, J. et al. (1997) *Science* 277: 91-

94]. Here, it has shown that *LEC1,* a critical activator of embryonic development, is similarly derepressed in pickle roots and in *pkl* seeds prior to germination.

The proposal that transcriptional inactivation of embryo-specific genes occurs after seed imbibition suggests that the final switch from embryonic to post-embryonic development occurs after seed maturation. This conclusion, in turn, suggests that seed-specific processes may be a developmental subset of embryo-specific processes, rather than a separate developmental program inserted between embryonic and post-embryonic developmental programs.

A general role for chromatin remodeling in GA signal transduction?

pkl plants exhibit numerous pleiotropies consistent with a defect in GA signal transduction. The rosette leaves are dark green with shortened petioles, time to flowering is increased, apical dominance is reduced, anther dehiscence is delayed, and *pkl* shoots accumulate bioactive GAs (Ogas, J. et al. (1997) *Science* 277: 91-94]; (manuscript in preparation). In addition, combining the pkl mutation with a *gai* mutation, which also perturbs GA signal transduction (Koorneef, M. et al. (1985) *Theor. Appl Genet* 61: 385-393; Talon, M. et al. (1990) *Planta* 182: 501-505; Wilson, R. N. and Somerville, C. (1992) *Plant Physiol.* 108: 495-502; Peng, J. and Harber, N. P. (1993) *Plant Cell* 5: 351-360; Wilson, R. N. and Somerville, C. (1995) *Plant Physiol.* 108: 495-502], gives rise to synergistic phenotypes [Ogas et al. (1997) *Science* 277: 91-94]. Based on these observations, it is proposed herein that *PKL* plays a general role in GA signal transduction. It is hypothesized in this study that GA promotes transitions from differentiation state A to differentiation state B by activating expression of genes necessary for state B and by repressing expression of genes necessary for state A via a PKL-dependent pathway. This model does not preclude the possibility that *PKL* activity may be stimulated by factors other than GA.

In conclusion, cloning of a gene necessary for repression of embryonic identity has lead to the proposition that a GA-modulated chromatin remodeling factor mediates a developmental transition in *Arabidopsis*. It is anticipated that further characterization of *PKL*, and identification of proteins that either regulate or are targets of *PKL*, will shed light on the mechanism of GA signal transduction and the role of GA in regulating differentiation and development in *Arabidopsis*. It remains to be determined whether CHD proteins in animal systems will play an analogous role in hormone-mediated developmental events.

EXAMPLE 4

Generation of Mutant PKL by a Dominant Negative Strategy

It has previously been demonstrated that a point mutation of a conserved lysine in the ATPase/helicase domain of SWI/SNF proteins generates a dominant negative mutant form of the protein [Chavari et al., (1993) *Nature* 366:170-174). By mutating the analogous mutation in PKL (by mutating Lys-304 to an Arg residue), a dominant negative version of PKL may be generated. This mutant allele of *PKL* may be generated by a PCR strategy.

A complementation construct for PKL was generated that includes the PKL cDNA flanked by 1.1 kb of upstream genomic sequence (to the BstBI site) and 1.4 kb of downstream genomic sequence (to the NcoI site). The construct was generated by performing overlap PCR on PKL cDNA with three DNA fragments: the genomic fragment upstream of the PKL start codon to the BstBI site, the PKL cDNA and the genomic fragment downstream of the termination codon to the NcoI site. A BstBI – XhoI fragment (2.1 kb) from this construct has been subcloned into a modified pBluescript vector (pJO674). The modified pBluescript vector pJ0674 was formed by ligating in a cassette generated by annealing the primers JOpr386 (5'-CTTCGAACTCGAGGGATCCCCATGGCTAGCAGCT-3'), shown in SEQ ID NO:26 (this is a synthetic sequence that includes "A"

followed by the recognition sequence of BstB1, Xhol, Bam HI, Ncol, Nhe I and sequence "AGCT" wherein the last "G" in the Ncol recognition sequence and the first "G" in the Nhel recognition sequence overlap) and JOpr387(5'-GCTAGCCATGGGGATCCCTCGAGTTCGAAGGTAC), as shown in SEQ ID NO:27 (this is a synthetic sequence complementary to SEQ ID NO:26) after pBluescript was cut with KpnI and SacI. The resulting cassette include the following restriction sites: BstB1, Xhol, Bam HI, Ncol and Nhel. 2 separate PCR reactions have been performed using this vector as a substrate. 1 PCR reaction uses a T3 primer with the following primer shown in SEQ ID NO:11 (JOpr516) 5'-GAAATGGGACTAGGCAGGACAATTCAAAGC-3' (nucleotides 895-924 in SEQ ID NO:1) where the underlined G is designed to replace an A residue in the wild-type PKL sequence and introduce the Lys-304 to Arg-304 mutation. This reaction generates a 272 bp fragment. The other PCR reaction uses a T7 primer with the following primer shown in SEQ ID NO:12 (JOpr517) 5'-GCTTTGAATTGTCCTGCCTAGTCCCATTTC-3' (sequence complementary to SEQ ID NO:1 from nucleotides 924-895) where the underlined C is designed to replace a T residue in the wild-type PKL sequence and introduce the Lys-304 to Arg-304 mutation. This reaction generates a 2094 bp fragment. Overlap PCR can then be done by adding the 272 bp and 2094 bp fragment together along with the T3 and T7 primers generating a 2.3 kb fragment. This fragment will be digested with BstBI and Xhol, cloned back into pJO674 and then sequenced to verify introduction of the mutation. This vector will then be cut with BstBI and Xhol and ligated into a pBluescript-based vector carrying the complementation construct (pJO765, formed by ligating the complementation fragment into pJO674 cut with BstBI and Ncol) cut with BstBI and Xhol, resulting in generation of a complementation construct that carries the dominant negative mutation. This construct will then be transferred to a binary vector [a modified pCAMBIA3300, pJO630, which is formed by digesting pCAMBIA3300 with BstXI and EcoRI and ligating in the

cassette generated by annealing primers JOpr232 (5'-CCAGGTACCTGG-3'), shown in SEQ ID NO:28 and JOpr233 (5'-AATTCCAGGTACCTGGCATG-3'), shown in SEQ ID NO:29] and transformed into wild-type plants to verify generation of a mutant *pkl* phenotype. These sequences are synthetic sequences that anneal to form a cassette that has ends that are compatible to BstXI and EcoRI digested pCAMBIA3300. The entire sequence of JOpr232 is a new site that when cut with BstXI generates ends that are compatible with KpnI ends. The cassette thus recreates a BstXI site with KpnI compatible ends. The PCR reactions and subcloning are performed as known in the art, and as described, for example, in Sambrook et al. (Eds.), *Molecular Cloning: A Laboratory Manua*l, 2nd ed., Cold Spring Harbor Laboratory Press (1989).

A conditional version of this dominant negative allele may be made by fusing the gene to the glucocorticoid receptor [Lloyd et al., (1994) *Science* 266:436-439). A clone of the rat glucocorticoid receptor (GR) was obtained from Alan Lloyd, at the University of Texas, Austin, Texas. The clone included SEQ ID NO:30 (5'-TCTAGAGGATCCTGAAGCTCGAAAAACAAAGAAAAAAA-3'), that is fused to nucleotides 1569-2407 of rat glucocorticoid receptor cDNA found in Genbank Accession No. Y12264. SEQ ID NO:30 was used to add spacers and restriction sites to the clone. A PCR reaction has been performed with this GR clone as a substrate and the following primers: JOpr533 (5'-AAGCCAAAGAACATGGTCGTTGATCTAGAGGATCCTGAAGCTCGAAA-3') shown in SEQ ID NO:13 (the first 24 nucleotides are nucleotides 4129-4152 of SEQ ID NO:1 whereas the last 23 nucleotides are nucleotides 2-24 of SEQ ID NO:30 of the rat glucocorticoid receptor cDNA found in Genbank Accession No. Y12264) and JOpr534 (5'-GAATCTTGATTTACCAGTTGAGTCATTTTTGATGAAACAGAAGCTTTTTGAT-3') (the first 25 nucleotides are nucleotides complementary to nucleotides 4153-4177 of SEQ ID NO:1 and the last 27 nucleotides are

complementary to nucleotides 2407-2381 of the glucocorticoid receptor cDNA found in Genbank Accession No. Y12264) shown in SEQ ID NO:14, which are designed to add PKL sequences to the end of the GR fragment such that overlap PCR can be performed. A BamHI –NcoI fragment of the complementation construct has been subcloned into pJO674, generating vector pJO724. pJO724 may be the substrate for 2 PCR reactions. One reaction can use the T7 primer and JOpr398 (5'-ATCAACGACCATGTTCTTTGG-3') (sequence complementary to nucleotides 4152-4132 of SEQ ID NO:1), shown in SEQ ID NO:15, generating a 883 bp fragment. The other reaction will use the T3 primer and JOpr401 (5'- TGACTCAACTGGTAAATCAAGA-3') (nucleotides 4153-4174 of SEQ ID NO:1), shown in SEQ ID NO:16, generating a 1.5 kb fragment. Overlap PCR can then be performed using 883 bp fragment and the GR fragment with the T7 primer and JOpr534. Overlap PCR can then be performed again using the product of this PCR reaction and the 1.5 kb fragment using the T7 primer and the T3 primer. This PCR product can then be digested with BamHI and NcoI and cloned back into pJO674 digested with the same. The construct will then be sequenced to verify identity. This construct can then be digested with BamHI and NcoI and ligated to the dominant-negative version of the complementation construct to generate a C-terminal fusion of GR to the mutant PKL protein. Once again, this construct can be transferred to a binary vector (pJO630) and transformed into wild-type plants to verify that a mutant *pkl* phenotype will be generated upon addition of dexamethasone.

If necessary, the dominant-negative version of the gene may be overexpressed in order to generate a phenotype. In this case, the mutated ORF (+/- GR) can be cloned downstream of a constitutive high level promoter such as the 35S promoter in a binary vector.

In all of Examples 4-6 described herein, ribonuclease protection assays will be performed to verify expression of the mutant transcript. The

pkl phenotype will be assayed by penetrance of the pickle root phenotype and by the rosette phenotype [Ogas, J. et al. (1997) *Science* 277:91-94].

EXAMPLE 5

Generation of Mutant PKL by Antisense Procedures

Two constructs for inhibiting expression of endogenous *PKL* by iRNA may be generated. These constructs are based on sequence comparison between PKL and PKR2, which is another CHD protein that exhibits high sequence similarity to PKL. A fragment of PKL may be cloned into the vector pRNA69, which results in formation of the following construct: 35S promoter – *PKL* frag in sense orientation – intron – the same PKL frag in antisense orientation – terminator. Vector pRNA69 is a bacterial vector that was obtained from John Bowman at UC Davis.

The sequence of the *PKL* cDNA that is being targeted in the first construct is from nucleotide 2 to nucleotide 361 in SEQ ID NO:1. This fragment was generated by performing PCR on PKL cDNA with the following primers: JOpr442 (5'-CCGCTCGAGTGAGTAGTTTGGTGGAGAGGC-3') found in SEQ ID NO:17 (the first 3 nucleotides are used as spacers so the restriction enzyme will cut properly, the next 6 nucleotides represent the XhoI recognition sequence and the last 21 nucleotides are nucleotides 2-22 of SEQ ID NO:1) and JOpr443 (5'-CCGGAATTCCATCGGAGGAACCTTGTTCAC-3'), found in SEQ ID NO:18 (the first 3 nucleotides are used as spacers so the restriction enzyme will cut properly, the next 6 nucleotides represent the Eco RI recognition sequence whereas the last 21 nucleotides are complementary to nucleotides 361-341 of SEQ ID NO:1), for the cloning the sense orientation (as a XhoI-EcoRI fragment) and JOpr444 (5'-CGCGGATCCCATCGGAGGAACCTTGTTCAC-3'), shown in SEQ ID NO:19 (the first 3 nucleotides are used as spacers so the restriction enzyme will cut properly, the next 6 nucleotides represent the Bam HI

recognition sequence and the last 21 nucleotides are complementary to nucleotides 361-341 of SEQ ID NO:1) and JOpr445 (5'-TGCTCTAGATGAGTAGTTTGGTGGAGAGGC-3'), shown in SEQ ID NO:20 (the first 3 nucleotides are used as spacers so the restriction enzyme will cut properly, the next 6 nucleotides represent the XbaI recognition sequence and the last 21 nucleotides are nucleotides 2-22 of SEQ ID NO:1), for cloning the antisense orientation (as a BamHI-XbaI fragment) into pRNA69.

The sequence of the *PKL* cDNA that is being targeted in the second construct is from nucleotide 3330 to nucleotide 3710 in SEQ ID NO:1. This fragment was generated by performing PCR on PKL cDNA with the following primers: JOpr446 (5'-CCGCTCGAGCCCTCACATAAGTTTGTCTGC-3'), shown in SEQ ID NO:21 (the first 3 nucleotides are used as spacers so the restriction enzyme will cut properly, the next 6 nucleotides represent the XhoI recognition sequence and the last 21 nucleotides are nucleotides 3330-3349 of SEQ ID NO:1), and JOpr447 (5'-CCGGAATTCGTCTTAGGAAGTCCATCAAGC-3'), shown in SEQ ID NO:22 (the first 3 nucleotides are used as spacers so the restriction enzyme will cut properly, the next 6 nucleotides represent the Eco RI recognition sequence and the last 21 nucleotides are complementary to nucleotides 3710-3690 of SEQ ID NO:1), for the cloning the sense orientation (as a XhoI-EcoRI fragment) and JOpr448 (5'-CGCGGATCCGTCTTAGGAAGTCCATCAAGC-3'), found in SEQ ID NO:23 (the first 3 nucleotides are used as spacers so the restriction enzyme will cut properly, the next 6 nucleotides represent the Bam HI recognition sequence whereas the last 21 bases are nucleotides 3330-3351 in SEQ ID NO:1), and JOpr449 (5'-TGCTCTAGACCCTCACATAAGTTTGTCTGC-3'), shown in SEQ ID NO:24 (the first 3 nucleotides are used as spacers so the restriction enzyme will cut properly, the next 6 nucleotides represent the XbaI recognition

sequence and the last 21 nucleotides are nucleotides 3330-3350 in SEQ ID NO:1)for cloning the antisense orientation (as a BamHI-XbaI fragment) into pRNA69.

The pRNA69 constructs may then be ligated into the binary vector pBART by making use of the flanking NotI sites. Wild-type plants may then be transformed by these constructs by vacuum infiltration. The plants may then be screened for a mutant pkl phenotype as described for Example 5.

EXAMPLE 6

Generation of Mutant PKL by Domain Deletion

It has been shown that removing the DNA-binding portion of *CHD1* in *S. cerevisiae* generates an inactive form of the protein [Woodage et al., (1997) *PNAS* 94:11472-11477). By specifically deleting the DNA-binding domain (aa 1069 - 1095) or any of the other domains, a dominant negative version of *PKL* may be produced. The XhoI-BamHI fragment of the PKL cDNA sequence has been cloned into pJO687, a vector obtained by introducing this fragment into a pJO674 vector formed as described in Example 4. In order to delete the putative DNA binding domain of PKL, PCR mutagenesis may be used. Briefly, a PCR reaction may be performed using pJO687 as a substrate and T7 and the oligo 5'-CGCGGATCCTTTTTCCACTTCTCAGTCCGGG-3', shown in SEQ ID NO:25 (the first 3 nucleotides are used as spacers so the restriction enzyme will cut properly, the next 6 nucleotides represent the Bam HI recognition sequence and the last 21 nucleotides are complementary to nucleotides 3202-3181 of SEQ ID NO:1), as a primer. The product can be digested with XhoI and BamHI and cloned into pJO674 cut with the same, and then can be sequenced to verify introduction of the mutation. This vector can then be cut with XhoI and BamHI and ligated into a pBluescript-based vector, carrying the complementation construct (pJO765) cut with the same, resulting in generation of a complementation construct that carries PKL deleted for the DNA binding domain. This construct can then

be transferred to a binary vector (a modified pCAMBIA3300, pJO630) formed as described in Example 4. Wild-type plants may then be transformed by methods described above with the vector to verify generation of a mutant pkl phenotype.

If necessary, the domain-deleted version of the gene can be overexpressed in order to generate a phenotype. If overexpression is desired, the mutated ORF can be cloned downstream of a constitutive high level promoter, such as the 35S promoter, in a binary vector.

While the invention has been illustrated and described in detail in the drawings and foregoing description, the same is to be considered as illustrative and not restrictive in character, it being understood that only the preferred embodiment has been shown and described and that all changes and modifications that come within the spirit of the invention are desired to be protected. In addition, all references cited herein are indicative of the level of skill in the art and are hereby incorporated by reference in their entirety.

CLAIMS

What is claimed is:

1. A method of transforming a host cell, comprising introducing into a host cell a nucleic acid molecule encoding a protein having at least one chromo domain, a helicase domain and a DNA binding domain, said protein expressed in an amount sufficient to regulate developmental identity.

2. The method of claim 1, wherein said nucleic acid molecule further encodes a protein having at least one zinc finger domain.

3. The method of claim 2, wherein said nucleic acid molecule further encodes a second chromo domain.

4. The method of claim 1, wherein said chromo domain is encoded by a nucleic acid molecule having a nucleotide sequence having at least about 50% identity to the nucleotide sequence set forth in SEQ ID NO:1 from nucleotide 343 to nucleotide 453, said helicase domain is encoded by a nucleic acid molecule having a nucleotide sequence having at least about 50% identity to the nucleotide sequence set forth in SEQ ID NO:1 from nucleotide 877 to nucleotide 2217, and said DNA binding domain is encoded by a nucleic acid molecule having a nucleotide sequence having at least about 50% identity to the nucleotide sequence set forth in SEQ ID NO:1 from nucleotide 3205 to nucleotide 3285.

5. The method of claim 2, wherein said zinc finger domain is encoded by a nucleic acid molecule having a nucleotide sequence having at least about 50% identity to the nucleotide sequence set forth in SEQ ID NO: 1 from nucleotide 145 to nucleotide 288.

6. The method of claim 3, wherein said second chromo domain is encoded by a nucleic acid molecule having a nucleotide sequence having at least about 50% identity to the nucleotide sequence set forth in SEQ ID NO:1 from nucleotide 571 to nucleotide 681.

7. The method of claim 1, wherein said nucleic acid molecule has a nucleotide sequence encoding protein domains selected from the group consisting of a chromo domain having an amino acid sequence having at least about 50% identity to the amino acid sequence set forth in SEQ ID NO:2 from amino acid 115 to amino acid 151, a helicase domain having an amino acid sequence having at least about 50% identity to the amino acid sequence set forth in SEQ ID NO:2 from amino acid 293 to amino acid 739 and a DNA binding domain having an amino acid sequence having at least about 50% identity to the amino acid sequence set forth in SEQ ID NO:2 from amino acid 1069 to amino acid 1095.

8. The method of claim 2, wherein said nucleic acid molecule has a nucleotide sequence encoding said zinc finger domain having an amino acid sequence having at least about 50% identity to the amino acid sequence set forth in SEQ ID NO:2 from amino acid 49 to amino acid 96.

9. The method of claim 3, wherein said nucleic acid molecule has a nucleotide sequence encoding said second chromo domain having an amino acid sequence having at least about 50% identity to the amino acid sequence set forth in SEQ ID NO:2 from amino acid 191 to amino acid 227.

10. The method of claim 1, wherein said host cell is a eukaryotic cell.

11. The method of claim 10, wherein said eukaryotic cell is a plant cell.

12. The method of claim 11, wherein said eukaryotic cell is an animal cell.

13. The method of claim 12, wherein said animal cell is a mammalian cell.

14. The method of claim 13, wherein said mammalian cell is a human cell.

15. The method of claim 1, further comprising deleting the nucleotide sequences encoding any one of said domains prior to said introducing.

16. The method of claim 1, wherein said protein has a point mutation in lysine 304 of SEQ ID NO:2.

17. The method of claim 16, wherein said mutation results in said lysine being replaced by an arginine.

18. The method of claim 1, wherein said protein encodes PKL.

19. The method of claim 18, wherein said PKL has an amino acid sequence as set forth in SEQ ID NO:2.

20. The method of claim 1, wherein said nucleic acid molecule has a nucleotide sequence having at least about 80% identity to the nucleotide sequence set forth in SEQ ID NO:1.

21. The method of claim 1, wherein said nucleic acid molecule further comprises a promoter operably linked to a terminal 5' end of said nucleotide sequence.

22. The method of claim 21, wherein said promoter is selected from the group consisting of a constitutive promoter, an inducible promoter and a cell-specific promoter.

23. The method of claim 21, wherein said promoter is a foreign promoter.

24. The method of claim 18, wherein said PKL functions in repressing embryonic identity in said plant.

25. The method of claim 1, wherein said nucleic acid molecule comprises a nucleotide sequence having substantial similarity to the nucleotide sequence set forth in SEQ ID NO:1.

26. A method of transforming a host cell, comprising introducing into a host cell a nucleic acid molecule encoding a protein functioning in regulating developmental identity, said protein having an amino acid sequence having at least about 50% identity to the amino acid sequence set forth in SEQ ID NO:2, said protein functioning in regulating developmental identity.

27. The method of claim 26, wherein said protein has an amino acid sequence having at least about 80% identity to the amino acid sequence set forth in SEQ ID NO:2.

28. The method of claim 27, wherein said protein has an amino acid sequence as set forth in SEQ ID NO:2.

29. The method of claim 26, wherein said nucleic acid molecule has a nucleotide sequence as set forth in SEQ ID NO:1.

30. The method of claim 26, wherein said host cell is a eukaryotic cell.

31. The method of claim 30, wherein said eukaryotic cell is a plant cell.

32. The method of claim 30, wherein said eukaryotic cell is an animal cell.

33. A method of transforming a host cell, comprising introducing into a host cell a nucleic acid molecule encoding a protein functioning in regulating developmental identity, said nucleic acid molecule having a nucleotide sequence having at least about 50% identity to the nucleotide sequence set forth in SEQ ID NO:1 from nucleotide 1 to nucleotide 4152.

34. The method of claim 33, wherein said protein functions in repressing embryonic identity.

35. The method of claim 33, wherein said nucleic acid molecule has a nucleotide sequence having at least about 80% identity to the nucleotide sequence set forth in SEQ ID NO:1 from nucleotide 1 to nucleotide 4152.

36. The method of claim 35, wherein said nucleic acid molecule has a nucleotide sequence as set forth in SEQ ID NO:1 from nucleotide 1 to nucleotide 4152.

37. The method of claim 33, wherein said nucleic acid molecule further comprises a promoter operably linked to a terminal 5' end of said nucleotide sequence.

38. A method of transforming a host cell, comprising:

(a) introducing into a host cell an antisense DNA or RNA molecule comprising a nucleotide sequence complementary to a length of nucleotides within a nucleic acid molecule encoding a protein having at least one chromo domain, a helicase domain and a DNA binding domain, said protein functioning in regulating developmental identity; and

(b) culturing said host cell under conditions effective for hybridization of said antisense molecule to nucleic acid of said host to regulate developmental identity.

39. The method of claim 38, wherein said protein encodes PKL.

40. The method of claim 38, wherein said nucleotide sequence is about 100 to about 1000 nucleotides in length.

41. The method of claim 38 wherein said nucleotide sequence is complementary to a region from about nucleotide 2 to about nucleotide 361 set forth in SEQ ID NO:1.

42. The method of claim 38, wherein said nucleotide sequence is complementary to a region from about nucleotide 3330 to about nucleotide 3710.

43. The method of claim 1, wherein said nucleic acid molecule further encodes a protein having at least one zinc finger domain.

44. The method of claim 38, wherein said nucleic acid molecule further encodes a second chromo domain.

45. A method of transforming a host cell, comprising:

(a) introducing into a host cell an antisense DNA or RNA molecule comprising a nucleotide sequence complementary to a length of nucleotides within a first nucleotide sequence having at least about 50% identity to the nucleotide sequence set forth in SEQ ID NO:1, said first nucleotide sequence encoding a protein functioning in regulating developmental identity; and

(b) culturing said host cell under conditions effective for hybridization of said antisense nucleotide sequence to nucleic acid of said host cell.

46. The method of claim 45, wherein said first nucleotide sequence has at least about 80% identity to the nucleotide sequence set forth in SEQ ID NO:1.

47. The method of claim 46, wherein said first nucleotide sequence is a nucleotide sequence that encodes PKL.

48. The method of claim 46, wherein said first nucleotide sequence is as set forth in SEQ ID NO:1 from nucleotide 1 to nucleotide 4152.

49. The method of claim 45, wherein said antisense molecule is about 100 to about 1000 nucleotides in length.

50. The method of claim 45, wherein said nucleotide sequence in complementary to a region from about nucleotide 2 to about nucleotide 361 set forth in SEQ ID NO:1.

51. The method of claim 45, wherein said nucleotide sequence is complementary to a region from about nucleotide 3330 to about nucleotide 3710 set forth in SEQ ID NO:1.

52. A method of transforming a host cell, comprising:

(a) introducing into a host cell a vector comprising a first nucleic acid molecule having a nucleotide sequence that is complementary to a nucleotide sequence having at least about 50% identity to a length of nucleotides within the nucleotide sequence set forth in SEQ ID NO:1, said nucleotide sequence encoding a protein functioning in regulating developmental identity;

(b) generating an antisense nucleic acid molecule complementary to an RNA transcript formed from SEQ ID NO:1; and

(b) culturing said host cell under conditions effective for hybridization of said antisense molecule to said RNA transcript of said host cell.

53. The method of claim 52, wherein said nucleic acid molecule has a nucleotide sequence that is complementary to a length of nucleotides within the nucleotide sequence set forth in SEQ ID NO:1.

54. The method of claim 52, wherein the antisense nucleic acid molecule is an RNA molecule.

55. A recombinant nucleic acid molecule, comprising:

(a) a nucleotide sequence encoding a protein functioning in regulating developmental identity, said protein having at least one chromo domain, a helicase domain and a DNA binding domain, said protein expressible in an amount sufficient to regulate developmental identity; and

(b) a foreign promoter operably linked to a terminal 5' end of said nucleotide sequence.

56. The method of claim 55, wherein said protein further has at least one zinc finger domain.

57. The method of claim 55, wherein said protein further has a second chromo domain.

58. A recombinant nucleic acid molecule, comprising:

(a) a nucleotide sequence encoding a protein functioning in regulating developmental identity, said protein having an amino acid sequence having at least about 50% identity to the amino acid sequence set forth in SEQ ID NO:2 ; and

(b) a foreign promoter operably linked to a terminal 5' end of said nucleotide sequence.

59. The molecule of claim 58, wherein said foreign promoter is selected from the group consisting of a constitutive promoter, an inducible promoter and a cell-specific promoter.

60. The molecule of claim 58, wherein said protein has an amino acid sequence having at least about 70% identity to the amino acid sequence set forth in SEQ ID NO:2.

61. The molecule of claim 58, wherein said protein has an amino acid sequence of PKL.

62. The molecule of claim 61, wherein said protein has an amino acid sequence as set forth in SEQ ID NO:2.

63. A recombinant nucleic acid molecule, comprising:

(a) a nucleotide sequence encoding a protein functioning in regulating developmental identity, said nucleotide sequence having at least about 50% identity to the nucleotide sequence set forth in SEQ ID NO:1 from nucleotide 1 to nucleotide 4152; and

(b) a foreign promoter operably linked to a terminal 5' end of said nucleotide sequence.

64. The molecule of claim 63, wherein said foreign promoter is selected from the group consisting of a constitutive promoter, an inducible promoter and a cell-specific promoter.

65. The molecule of claim 63, wherein said nucleotide sequence has at least about 80% identity to the nucleotide sequence set forth in SEQ ID NO:1 from nucleotide 1 to nucleotide 4152.

66. The molecule of claim 65, wherein said nucleotide sequence is as set forth in SEQ ID NO:1 from nucleotide 1 to nucleotide 4152.

67. An isolated nucleic acid molecule, comprising a nucleotide sequence encoding a protein functioning in regulating developmental identity, said nucleotide sequence encoding a protein having at least one chromo domain, a helicase domain and a DNA binding domain.

68. The molecule of claim 67, wherein said nucleic acid molecule further encodes a protein having at least one zinc finger domain.

69. The molecule of claim 68, wherein said nucleic acid molecule further encodes a second chromo domain.

70. A eukaryotic cell, comprising:

(a) an introduced nucleic acid molecule having a nucleotide sequence encoding a protein functioning in regulating developmental identity, said protein having an amino acid sequence having at least about 50% identity to the amino acid sequence set forth in SEQ ID NO:2.

71. The cell of claim 70, wherein said protein has an amino acid sequence having at least about 50% identity to the amino acid sequence set forth in SEQ ID NO:2.

72. The cell of claim 71, wherein said protein has an amino acid sequence as set forth in SEQ ID NO:2.

73. The cell of claim 70, wherein said cell is a plant cell.

74. The cell of claim 70, wherein said cell is an animal cell.

75. A transgenic plant, comprising:

(a) an introduced nucleic acid molecule having a nucleotide sequence encoding a plant protein functioning in regulating developmental identity, said protein having an amino acid sequence having at least about 50% identity to the amino acid sequence set forth in SEQ ID NO:2; and

(b) a foreign promoter operably linked to a terminal 5' end of said nucleotide sequence.

76. The transgenic plant of claim 75, wherein said nucleotide sequence is an antisense DNA or RNA molecule.

77. The transgenic plant of claim 75, wherein said protein has an amino acid sequence having at least about 80% identity to the amino acid sequence set forth in SEQ ID NO:2.

78. The transgenic plant of claim 77, wherein said protein has the amino acid sequence of PKL.

79. The transgenic plant of claim 78, wherein said amino acid sequence is as set forth in SEQ ID NO:2.

80. A recombinant protein, comprising a protein having an amino acid sequence having at least about 50% identity to the amino acid sequence set forth in SEQ ID NO:2.

81. The protein of claim 80, wherein said protein has an amino acid sequence having at least about 80% identity to the amino acid sequence set forth in SEQ ID NO:2.

82. The protein of claim 81, wherein said protein has an amino acid sequence as set forth in SEQ ID NO:2.

83. A method of producing a PKL protein, comprising:

(a) introducing a nucleotide sequence encoding a protein having at least about 50% identity to the amino acid sequence set forth in SEQ ID NO:2; and

(b) culturing said host cell under conditions effective to achieve expression of the PKL polypeptide.

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
1 March 2001 (01.03.2001)

PCT

(10) International Publication Number
WO 01/14519 A2

(51) International Patent Classification[7]: C12N

(21) International Application Number: PCT/US00/22725

(22) International Filing Date: 18 August 2000 (18.08.2000)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/149,975 20 August 1999 (20.08.1999) US

(71) Applicant *(for all designated States except US)*: **PURDUE RESEARCH FOUNDATION** [US/US]; Office of Technology Commercialization, 1291 Cumberland Avenue, West Lafayette, IN 47906 (US).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: **OGAS, Joseph, P.** [US/US]; 805 N. Chauncey Avenue, West Lafayette, IN 47906 (US). **SOMERVILLE, Chris, R.** [US/US]; 5 Valley Oak, Portola Valley, CA 94028 (US).

(74) Agents: **SCHWARTZ, Jason, J.** et al.; Woodard, Emhardt, Naughton, Moriarty & McNett, Bank One Center/Tower, Suite 3700, 111 Monument Circle, Indianapolis, IN 46204 (US).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published:
— *Without international search report and to be republished upon receipt of that report.*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

WO 01/14519 A2

(54) Title: METHODS AND COMPOSITIONS FOR REGULATING DEVELOPMENTAL IDENTITY

(57) Abstract: Purified PKL proteins that function in regulating developmental identity in host cells are provided. Nucleotide sequences encoding functional PKL proteins are also provided. The invention also provides recombinant vectors including the nucleotide sequences encoding PKL, eukaryotic host cells and transgenic plants that include the introduced nucleotide sequences described herein, and methods of transforming plants utilizing the constructs described herein.

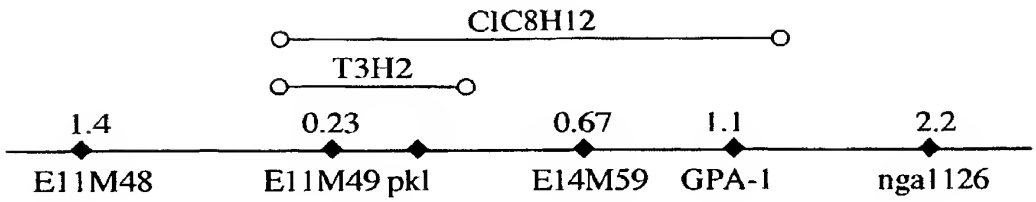


Fig. 1

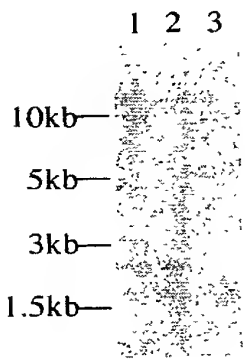


Fig. 2

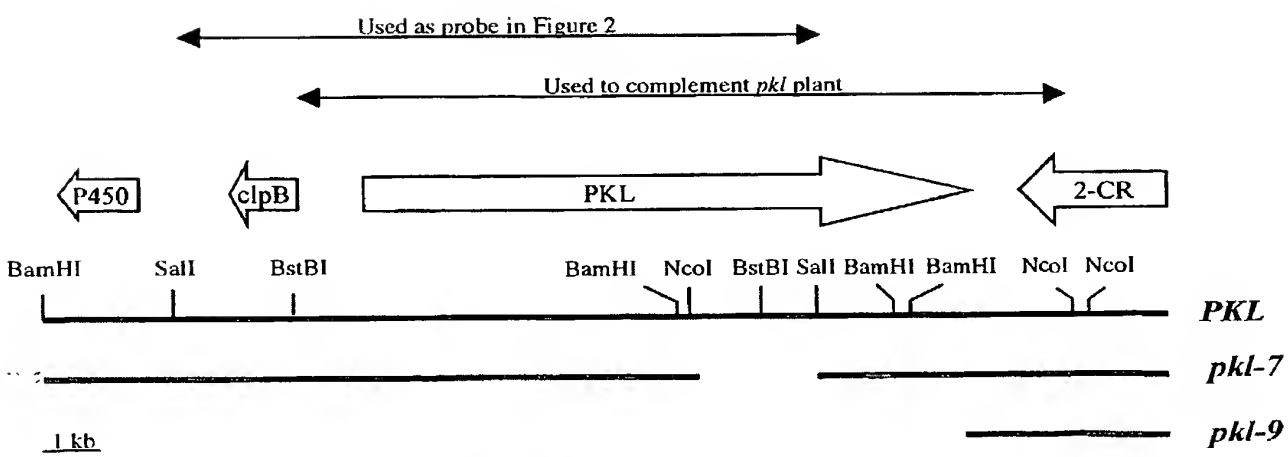


Fig. 3

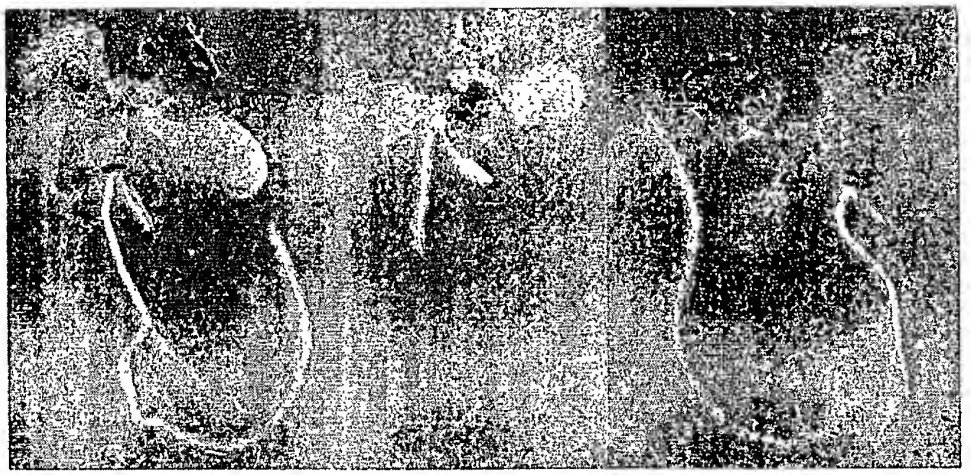


Fig. 4A

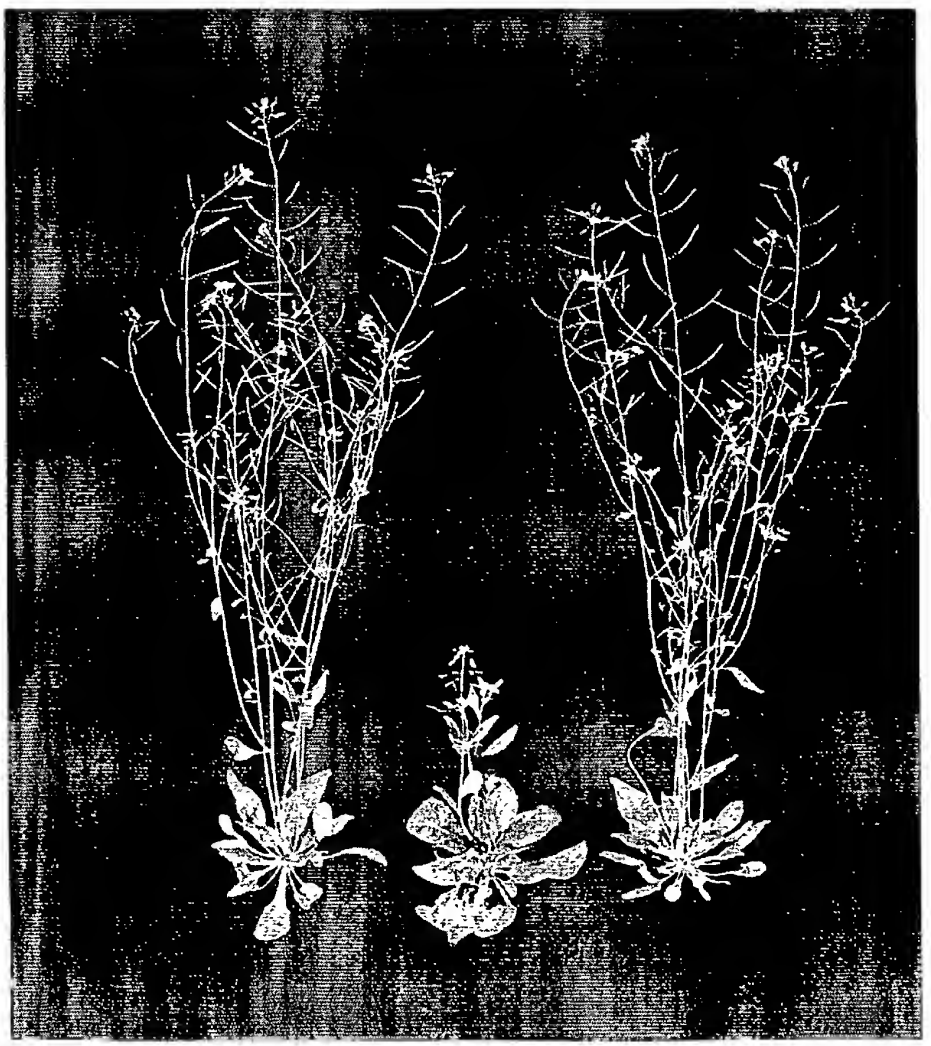


Fig. 4B

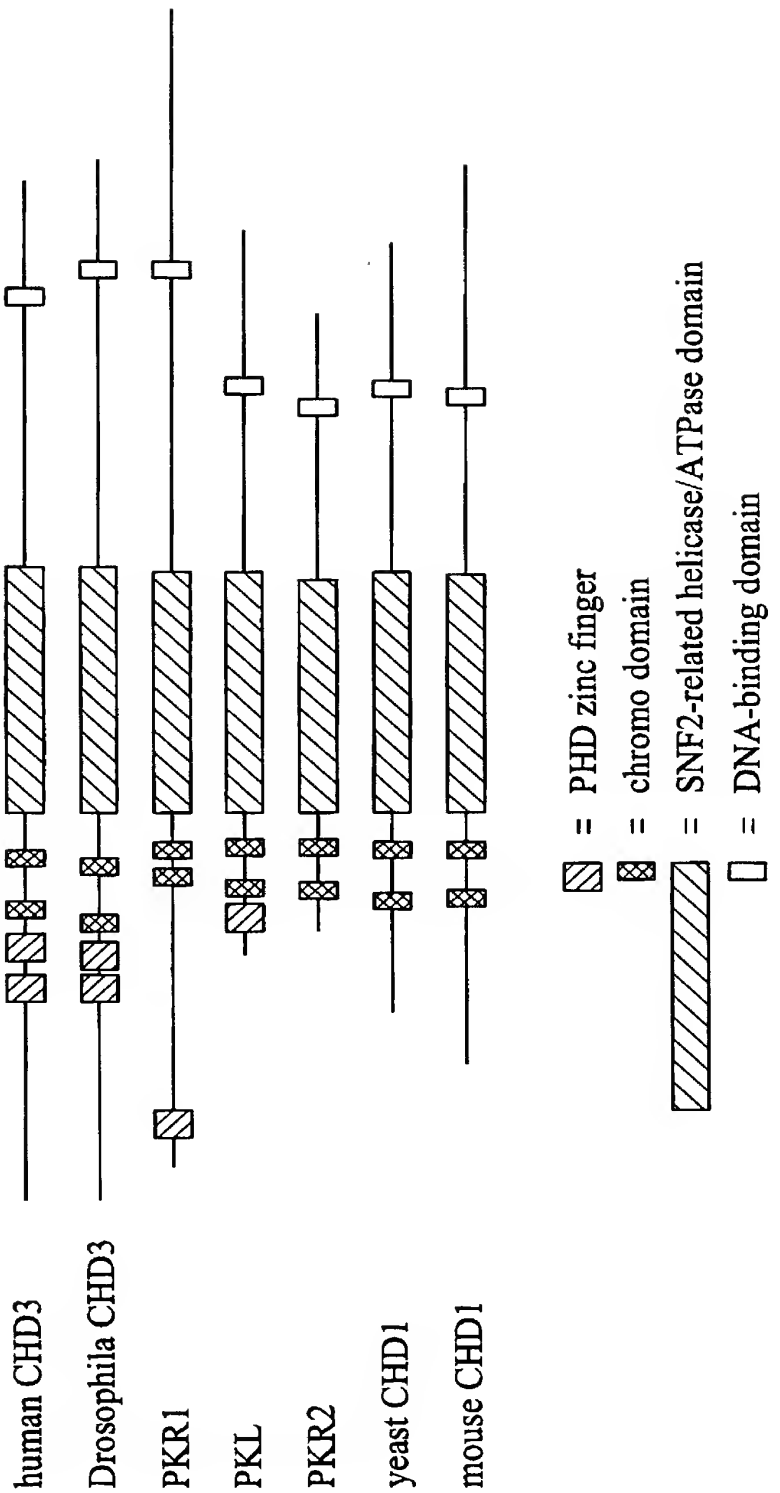

human CHD3
Drosophila CHD3
PKR1
PKL
PKR2
yeast CHD1
mouse CHD1
= PHD zinc finger
= chromo domain
= SNF2-related helicase/ATPase domain
= DNA-binding domain

Fig. 5

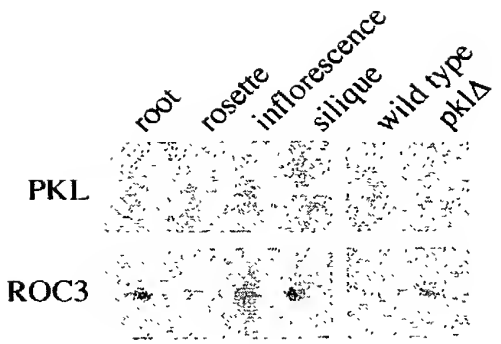


Fig. 6

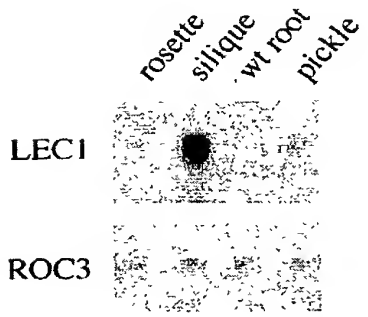


Fig. 7

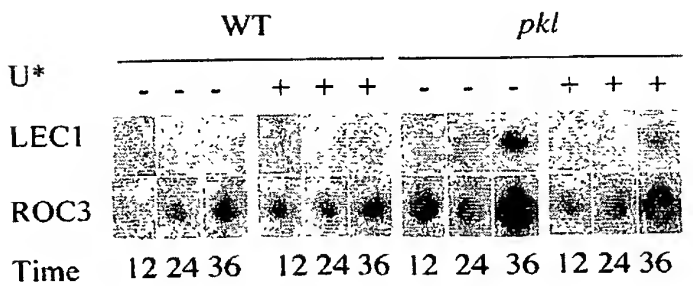


Fig. 8

PTO/SB/01 (10-01)
Approved for use through 10/31/2002 OMB 0651-0032
U.S. Patent and Trademark Office, U S. DEPARTMENT OF COMMERCE

DECLARATION FOR UTILITY OR DESIGN PATENT APPLICATION (37 CFR 1.63)

☐ Declaration Submitted with Initial Filing **OR** ☒ Declaration Submitted after Initial Filing (surcharge (37 CFR 1.16 (e)) required)

Attorney Docket Number	7024-509
First Named Inventor	Joseph P. OGAS
COMPLETE IF KNOWN	
Application Number	/
Filing Date	
Art Unit	
Examiner Name	

As the below named inventor, I hereby declare that:

My residence, mailing address, and citizenship are as stated below next to my name.

I believe I am the original and first inventor of the subject matter which is claimed and for which a patent is sought on the invention entitled:

METHODS AND COMPOSITIONS FOR REGULATING DEVELOPMENTAL IDENTITY

(Title of the Invention)

the specification of which

☐ is attached hereto

OR

☒ was filed on (MM/DD/YYYY) 18 August 2000 (18.08.00) as United States Application Number or PCT International Application Number PCT/US00/22725 and was amended on (MM/DD/YYYY) 08/28/2001 (if applicable)

I hereby state that I have reviewed and understand the contents of the above identified specification, including the claims, as amended by any amendment specifically referred to above

I acknowledge the duty to disclose information which is material to patentability as defined in 37 CFR 1 56, including for continuation-in-part applications, material information which became available between the filing date of the prior application and the national or PCT international filing date of the continuation-in-part application.

I hereby claim foreign priority benefits under 35 U.S.C. 119(a)-(d) or (f), or 365(b) of any foreign application(s) for patent, inventor's or plant breeder's rights certificate(s), or 365(a) of any PCT international application which designated at least one country other than the United States of America, listed below and have also identified below, by checking the box, any foreign application for patent, inventor's or plant breeder's rights certificate(s), or any PCT international application having a filing date before that of the application on which priority is claimed.

Prior Foreign Application Number(s)	Country	Foreign Filing Date (MM/DD/YYYY)	Priority Not Claimed	Certified Copy Attached? YES	Certified Copy Attached? NO
PCT/US00/22725	PCT	08/18/2000	☐	☐	☒
			☐	☐	☐
			☐	☐	☐
			☐	☐	☐

☐ Additional foreign application numbers are listed on a supplemental priority data sheet PTO/SB/02B attached hereto:

PTO/SB/01 (10-01)
Approved for use through 10/31/2002 OMB 0651-0032
U S Patent and Trademark Office, U S. DEPARTMENT OF COMMERCE

DECLARATION — Utility or Design Patent Application

Direct all correspondence to: ☐ Customer Number or Bar Code Label [ ] OR [X] Correspondence address below

Name: Gregory B. COY @ Woodard, Emhardt, Naughton, Moriarty & McNett

Address: Bank One Center/Tower, Suite 3700
111 Monument Circle

City: Indianapolis | State: IN | ZIP: 46203

Country: US | Telephone: 317-634-3456 | Fax: 317-637-7561

I hereby declare that all statements made herein of my own knowledge are true and that all statements made on information and belief are believed to be true; and further that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under 18 U.S C. 1001 and that such willful false statements may jeopardize the validity of the application or any patent issued thereon.

NAME OF SOLE OR FIRST INVENTOR : ☐ A petition has been filed for this unsigned inventor

Given Name (first and middle [if any]): JOSEPH P. | Family Name or Surname: OGAS

Inventor's Signature: [signature] | Date: 4/23/02

Residence: City West Lafayette | State IN | Country US | Citizenship US

Mailing Address: 805 N. Chauncey Avenue

City: West Lafayette | State: IN | ZIP: 47906 | Country: US

NAME OF SECOND INVENTOR: ☐ A petition has been filed for this unsigned inventor

Given Name (first and middle [if any]): CHRIS R. | Family Name or Surname: SOMERVILLE

Inventor's Signature: | Date:

Residence: City Portola Valley | State CA | Country US | Citizenship US

Mailing Address: 5 Valley Oak

City: Portola Valley | State: CA | ZIP: 94028 | Country: US

☐ Additional inventors are being named on the ____supplemental Additional Inventor(s) sheet(s) PTO/SB/02A attached hereto.

PTO/SB/01 (10-01)
Approved for use through 10/31/2002 OMB 0651-0032
U S Patent and Trademark Office, U S DEPARTMENT OF COMMERCE
Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it contains a valid OMB control number

DECLARATION — Utility or Design Patent Application

Direct all correspondence to. ☐ Customer Number or Bar Code Label [ ] OR ☒ Correspondence address below

Name Gregory B. COY @ Woodard, Emhardt, Naughton, Moriarty & McNett

Address Bank One Center/Tower, Suite 3700
111 Monument Circle

City	**State**	**ZIP**
Indianapolis	IN	46203

Country	**Telephone**	**Fax**
US	317-634-3456	317-637-7561

I hereby declare that all statements made herein of my own knowledge are true and that all statements made on information and belief are believed to be true; and further that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under 18 U.S C. 1001 and that such willful false statements may jeopardize the validity of the application or any patent issued thereon.

NAME OF SOLE OR FIRST INVENTOR : ☐ A petition has been filed for this unsigned inventor

Given Name (first and middle [if any])	**Family Name or Surname**
JOSEPH P.	OGAS

Inventor's Signature _____ **Date** _____

Residence: City	**State**	**Country**	**Citizenship**
West Lafayette	IN	US	US

Mailing Address 805 N. Chauncey Avenue

City	**State**	**ZIP**	**Country**
West Lafayette	IN	47906	US

NAME OF SECOND INVENTOR: ☐ A petition has been filed for this unsigned inventor

Given Name (first and middle [if any])	**Family Name or Surname**
CHRIS R.	SOMERVILLE

Inventor's Signature Chris R Somerville **Date** 04-24-02

Residence: City	**State**	**Country**	**Citizenship**
Portola Valley	CA	US	US

Mailing Address 5 Valley Oak

City	**State**	**ZIP**	**Country**
Portola Valley	CA	94028	US

☐ Additional inventors are being named on the ____supplemental Additional Inventor(s) sheet(s) PTO/SB/02A attached hereto

Please type a plus sign (+) inside this box → [+]

PTO/SB/02C (3-97)
Approved for use through 9/30/98 OMB 0651-0032
Patent and Trademark Office, U.S. DEPARTMENT OF COMMERCE

Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it contains a valid OMB control number

DECLARATION

REGISTERED PRACTITIONER INFORMATION (Supplemental Sheet)

Name	Registration Number	Name	Registration Number
Harold R. Woodard	#16,214		
C. David Emhardt	#18,483		
Joseph A. Naughton, Jr.	#19,814		
John V. Moriarty	#26,207		
John C. McNett	#25,533		
Thomas Q. Henry	#28,309		
James M. Durlacher	#28,840		
Charles R. Reeves	#28,750		
Vincent O. Wagner	#29,596		
Steve Zlatos	#30,123		
Spiro Bereveskos	#30,821		
Clifford W. Browning	#32,201		
R. Randall Frisk	#32,221		
Daniel J. Lueders	#32,581		
Kenneth A. Gandy	#33,386		
Timothy N. Thomas	#35,714		
Kurt N. Jones	#37,996		
John H. Allie	#39,088		
Holiday W. Banta	#40,311		
Troy J. Cole	#35,102		
L. Scott Paynter	#39,797		
Charles J. Meyer	#41,996		
Matthew R. Schantz	#40,800		
Gregory B. Coy	#40,967		
Lisa A. Hiday	#40,036		
John V. Daniluck	#40,581		
Christopher A. Brown	#41,642		
C. John Brannon	#44,557		
Arthur J. Usher, IV	#41,359		
Douglas A. Collier	#43,556		
Brad A. Schepers	#45,431		
James B. Myers	#42,021		
Scott J. Stevens	#29,446		
John M. Bradshaw	#46,573		
Charles P. Schmal	#45,082		
Edward E. Sowers	#36,015		
Quentin G. Cantrell	#47,469		

37

Burden Hour Statement: This form is estimated to take 0.4 hours to complete. Time will vary depending upon the needs of the individual case. Any comments on the amount of time you are required to complete this form should be sent to the Chief Information Officer, Patent and Trademark Office, Washington, DC 20231 DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS SEND TO Assistant Commissioner for Patents, Washington, DC 20231

Pklseq1.app
SEQUENCE LISTING

<110> Ogas, Joseph P.
Somerville, Christopher R.

<120> Methods and Compositions for Regulating Developmental Identity

<130> 7024-473

<140> Unknown

<141> 2000-08-18

<150> US 60/149,975

<151> 1999-08-20

<160> 30

<170> MS Notebook

<210> 1

<211> 4177

<212> DNA

<213> Arabidopsis thaliana

<220>

<400> 1

```
atg agt agt ttg gtg gag agg ctt cgc ata cga tct gat agg aaa cca   48

Met Ser Ser Leu Val Glu Arg Leu Arg Ile Arg Ser Asp Arg Lys Pro
  1               5                  10                  15

gtt tat aac cta gat gat tct gat gat gac gac ttc gtt cct aaa aaa   96

Val Tyr Asn Leu Asp Asp Ser Asp Asp Asp Asp Phe Val Pro Lys Lys
            20                  25                  30

gat cga acc ttt gag caa gtc gag gct att gtc aga act gat gcg aaa  144

Asp Arg Thr Phe Glu Gln Val Glu Ala Ile Val Arg Thr Asp Ala Lys
        35                  40                  45

gaa aat gca tgt cag gct tgt ggg gaa agt act aat ctt gta agc tgc  192
```

Pklseq1.app

```
Glu Asn Ala Cys Gln Ala Cys Gly Glu Ser Thr Asn Leu Val Ser Cys
 50                  55                  60

aat aca tgc act tat gcg ttc cat gct aaa tgc tta gtt cca cct ctt   240

Asn Thr Cys Thr Tyr Ala Phe His Ala Lys Cys Leu Val Pro Pro Leu
 65                  70                  75                  80

aaa gat gct tcc gtg gaa aat tgg aga tgc cct gaa tgt gtt agt cct   288

Lys Asp Ala Ser Val Glu Asn Trp Arg Cys Pro Glu Cys Val Ser Pro
                 85                  90                  95

ctt aac gag ata gat aag ata ttg gat tgt gaa atg cgt cct aca aaa   336

Leu Asn Glu Ile Asp Lys Ile Leu Asp Cys Glu Met Arg Pro Thr Lys
            100                 105                 110

tct agt gaa caa ggt tcc tcc gat gcg gaa ccg aag cca att ttt gtg   384

Ser Ser Glu Gln Gly Ser Ser Asp Ala Glu Pro Lys Pro Ile Phe Val
        115                 120                 125

aaa cag tat ctc gtg aag tgg aag gga tta tca tac ctt cac tgc tct   432

Lys Gln Tyr Leu Val Lys Trp Lys Gly Leu Ser Tyr Leu His Cys Ser
    130                 135                 140

tgg gtg cct gag aag gag ttc cag aag gct tat aag tca aat cat cgt   480

Trp Val Pro Glu Lys Glu Phe Gln Lys Ala Tyr Lys Ser Asn His Arg
145                 150                 155                 160

tta aaa acc aga gtg aac aat ttt cac cgt caa atg gag tca ttc aat   528

Leu Lys Thr Arg Val Asn Asn Phe His Arg Gln Met Glu Ser Phe Asn
                165                 170                 175

aac agc gaa gat gat ttt gtt gcc ata cgt cct gag tgg acc act gtt   576

Asn Ser Glu Asp Asp Phe Val Ala Ile Arg Pro Glu Trp Thr Thr Val
            180                 185                 190

gat cgg att ctt gcc tgc aga gag gaa gat ggg gag ctg gaa tat ctt   624

Asp Arg Ile Leu Ala Cys Arg Glu Glu Asp Gly Glu Leu Glu Tyr Leu
        195                 200                 205

gtc aaa tat aaa gag cta tcc tat gat gaa tgt tat tgg gag tca gaa   672

Val Lys Tyr Lys Glu Leu Ser Tyr Asp Glu Cys Tyr Trp Glu Ser Glu
```

```
    210                 215                 220

tca gac atc tca acc ttc cag aat gaa att caa agg ttc aag gat gta   720

Ser Asp Ile Ser Thr Phe Gln Asn Glu Ile Gln Arg Phe Lys Asp Val
225                 230                 235                 240

aat tct aga act cgc aga agt aaa gat gtt gac cat aaa aga aat ccc   768

Asn Ser Arg Thr Arg Arg Ser Lys Asp Val Asp His Lys Arg Asn Pro
                245                 250                 255

aga gac ttt caa cag ttt gat cat act cct gaa ttc ctc aaa ggc ttg   816

Arg Asp Phe Gln Gln Phe Asp His Thr Pro Glu Phe Leu Lys Gly Leu
            260                 265                 270

tta cat cca tac cag ctt gag gga ctt aat ttt ttg cgg ttc tcg tgg   864

Leu His Pro Tyr Gln Leu Glu Gly Leu Asn Phe Leu Arg Phe Ser Trp
        275                 280                 285

tca aaa cag acg cat gta atc ctt gct gat gaa atg gga cta ggc aag   912

Ser Lys Gln Thr His Val Ile Leu Ala Asp Glu Met Gly Leu Gly Lys
    290                 295                 300

aca att caa agc att gcc ctt tta gct tca ctt ttt gag gag aac ctc   960

Thr Ile Gln Ser Ile Ala Leu Leu Ala Ser Leu Phe Glu Glu Asn Leu
305                 310                 315                 320

att ccg cat ttg gta att gct cct cta tcg act ctg cgt aac tgg gag   1008

Ile Pro His Leu Val Ile Ala Pro Leu Ser Thr Leu Arg Asn Trp Glu
                325                 330                 335

aga gag ttt gcc aca tgg gcc cca cag atg aac gtg gtt atg tat ttt   1056

Arg Glu Phe Ala Thr Trp Ala Pro Gln Met Asn Val Val Met Tyr Phe
            340                 345                 350

ggc act gcg caa gct cga gca gtt atc aga gaa cat gag ttt tac tta   1104

Gly Thr Ala Gln Ala Arg Ala Val Ile Arg Glu His Glu Phe Tyr Leu
        355                 360                 365

tcg aaa gat caa aaa aag atc aag aaa aag aaa tct gga caa ata agt   1152

Ser Lys Asp Gln Lys Lys Ile Lys Lys Lys Lys Ser Gly Gln Ile Ser
    370                 375                 380
```

```
agc gaa agc aag caa aaa aga atc aag ttt gat gtc ctc ctc aca tcg   1200

Ser Glu Ser Lys Gln Lys Arg Ile Lys Phe Asp Val Leu Leu Thr Ser
385                 390                 395                 400

tat gag atg atc aac cta gat tca gca gtt cta aaa cca att aag tgg   1248

Tyr Glu Met Ile Asn Leu Asp Ser Ala Val Leu Lys Pro Ile Lys Trp
                405                 410                 415

gag tgc atg att gtt gat gaa ggt cat cga ctg aaa aat aag gat tca   1296

Glu Cys Met Ile Val Asp Glu Gly His Arg Leu Lys Asn Lys Asp Ser
            420                 425                 430

aag ctg ttc tct tca ttg aca cag tat tca agt aac cac cgt att ctt   1344

Lys Leu Phe Ser Ser Leu Thr Gln Tyr Ser Ser Asn His Arg Ile Leu
        435                 440                 445

ctg aca gga aca cca ctt cag aac aac ttg gat gaa ctt ttc atg ctc   1392

Leu Thr Gly Thr Pro Leu Gln Asn Asn Leu Asp Glu Leu Phe Met Leu
    450                 455                 460

atg cat ttt ctt gat gcg ggg aag ttt gga agt ttg gag gag ttc cag   1440

Met His Phe Leu Asp Ala Gly Lys Phe Gly Ser Leu Glu Glu Phe Gln
465                 470                 475                 480

gag gag ttc aaa gat att aat caa gag gag cag atc tca agg ttg cac   1488

Glu Glu Phe Lys Asp Ile Asn Gln Glu Glu Gln Ile Ser Arg Leu His
                485                 490                 495

aaa atg ttg gct cca cat ttg ctc aga agg gta aaa aaa gac gta atg   1536

Lys Met Leu Ala Pro His Leu Leu Arg Arg Val Lys Lys Asp Val Met
            500                 505                 510

aaa gac atg ccc ccc aaa aag gag ctc att ttg cgt gtt gat ctg agc   1584

Lys Asp Met Pro Pro Lys Lys Glu Leu Ile Leu Arg Val Asp Leu Ser
        515                 520                 525

agt ctg cag aaa gaa tat tac aaa gct att ttt acc cgt aat tat caa   1632

Ser Leu Gln Lys Glu Tyr Tyr Lys Ala Ile Phe Thr Arg Asn Tyr Gln
    530                 535                 540
```

```
gta ttg aca aaa aag gga ggt gct caa att tcc ctt aat aac att atg   1680

Val Leu Thr Lys Lys Gly Gly Ala Gln Ile Ser Leu Asn Asn Ile Met
545                 550                 555                 560

atg gaa tta cga aaa gta tgc tgc cat cct tat atg cta gag ggt gtt   1728

Met Glu Leu Arg Lys Val Cys Cys His Pro Tyr Met Leu Glu Gly Val
                565                 570                 575

gag cca gtt att cac gac gca aat gaa gct ttc aaa caa ctt ttg gag   1776

Glu Pro Val Ile His Asp Ala Asn Glu Ala Phe Lys Gln Leu Leu Glu
            580                 585                 590

tct tgt gga aag ctg caa ctt cta gat aaa atg atg gtc aaa ctg aaa   1824

Ser Cys Gly Lys Leu Gln Leu Leu Asp Lys Met Met Val Lys Leu Lys
        595                 600                 605

gag caa gga cac aga gtc cta ata tac aca cag ttt cag cat atg ctg   1872

Glu Gln Gly His Arg Val Leu Ile Tyr Thr Gln Phe Gln His Met Leu
    610                 615                 620

gac tta ctt gaa gac tac tgt acc cat aag aaa tgg cag tac gag cga   1920

Asp Leu Leu Glu Asp Tyr Cys Thr His Lys Lys Trp Gln Tyr Glu Arg
625                 630                 635                 640

att gat gga aag gtt ggc gga gct gag cgg caa ata cgc ata gat cgg   1968

Ile Asp Gly Lys Val Gly Gly Ala Glu Arg Gln Ile Arg Ile Asp Arg
                645                 650                 655

ttc aat gcc aaa aat tct aac aag ttt tgt ttt ttg ctc tcc aca aga   2016

Phe Asn Ala Lys Asn Ser Asn Lys Phe Cys Phe Leu Leu Ser Thr Arg
            660                 665                 670

gct ggt ggc tta gga ata aat ctt gca acg gct gat aca gta atc att   2064

Ala Gly Gly Leu Gly Ile Asn Leu Ala Thr Ala Asp Thr Val Ile Ile
        675                 680                 685

tat gac agt gac tgg aat cct cat gct gat ctt caa gca atg gct aga   2112

Tyr Asp Ser Asp Trp Asn Pro His Ala Asp Leu Gln Ala Met Ala Arg
    690                 695                 700

gct cat cga ctt ggc caa aca aat aag gtg atg att tat agg ctc ata   2160
```

```
Ala His Arg Leu Gly Gln Thr Asn Lys Val Met Ile Tyr Arg Leu Ile
705                 710                 715                 720

aac cga ggc acc att gaa gaa agg atg atg caa ttg act aaa aag aaa   2208

Asn Arg Gly Thr Ile Glu Glu Arg Met Met Gln Leu Thr Lys Lys Lys
                725                 730                 735

atg gtt cta gag cat ctt gtt gtt ggg aaa ctc aaa aca caa aac att   2256

Met Val Leu Glu His Leu Val Val Gly Lys Leu Lys Thr Gln Asn Ile
            740                 745                 750

aat cag gaa gag tta gat gac atc atc agg tat gga tca aag gag ctt   2304

Asn Gln Glu Glu Leu Asp Asp Ile Ile Arg Tyr Gly Ser Lys Glu Leu
        755                 760                 765

ttt gct agt gaa gat gat gaa gca gga aag tct gga aaa att cat tat   2352

Phe Ala Ser Glu Asp Asp Glu Ala Gly Lys Ser Gly Lys Ile His Tyr
    770                 775                 780

gat gat gcg gct ata gac aaa ttg ctt gat cgt gat ctc gtg gag gca   2400

Asp Asp Ala Ala Ile Asp Lys Leu Leu Asp Arg Asp Leu Val Glu Ala
785                 790                 795                 800

gag gaa gtc tca gtg gat gat gaa gag gag aat gga ttc tta aag gct   2448

Glu Glu Val Ser Val Asp Asp Glu Glu Glu Asn Gly Phe Leu Lys Ala
                805                 810                 815

ttc aag gtg gct aat ttt gaa tat ata gat gaa aat gag gca gca gca   2496

Phe Lys Val Ala Asn Phe Glu Tyr Ile Asp Glu Asn Glu Ala Ala Ala
            820                 825                 830

tta gag gca cag aga gtc gct gct gaa agc aaa tct tca gca ggc aat   2544

Leu Glu Ala Gln Arg Val Ala Ala Glu Ser Lys Ser Ser Ala Gly Asn
        835                 840                 845

tct gat aga gca agt tat tgg gaa gag ttg tta aaa gat aaa ttt gag   2592

Ser Asp Arg Ala Ser Tyr Trp Glu Glu Leu Leu Lys Asp Lys Phe Glu
    850                 855                 860

ctg cac cag gct gag gag ctt aat gct ctt gga aaa agg aag aga agt   2640
```

```
Leu His Gln Ala Glu Glu Leu Asn Ala Leu Gly Lys Arg Lys Arg Ser
865                 870                 875                 880

cgc aag cag ttg gta tcc att gaa gaa gat gat ctt gct ggt ttg gaa   2688

Arg Lys Gln Leu Val Ser Ile Glu Glu Asp Asp Leu Ala Gly Leu Glu
                885                 890                 895

gat gtg agc tct gat gga gat gaa agt tat gaa gct gag tca aca gat   2736

Asp Val Ser Ser Asp Gly Asp Glu Ser Tyr Glu Ala Glu Ser Thr Asp
            900                 905                 910

ggt gaa gca gca gga caa gga gtt cag acg ggt cga cgg ccg tac aga   2784

Gly Glu Ala Ala Gly Gln Gly Val Gln Thr Gly Arg Arg Pro Tyr Arg
        915                 920                 925

aga aag ggt cgc gat aat ttg gaa cca act ccg ttg atg gaa ggt gag   2832

Arg Lys Gly Arg Asp Asn Leu Glu Pro Thr Pro Leu Met Glu Gly Glu
    930                 935                 940

ggg aga tct ttc aga gta ctg ggt ttc aac cag agt caa agg gcc att   2880

Gly Arg Ser Phe Arg Val Leu Gly Phe Asn Gln Ser Gln Arg Ala Ile
945                 950                 955                 960

ttt gta cag act ttg atg agg tat gga gct ggc aat ttt gat tgg aag   2928

Phe Val Gln Thr Leu Met Arg Tyr Gly Ala Gly Asn Phe Asp Trp Lys
                965                 970                 975

gag ttt gtt cct cgc tta aag cag aag acc ttt gaa gaa ata aat gaa   2976

Glu Phe Val Pro Arg Leu Lys Gln Lys Thr Phe Glu Glu Ile Asn Glu
            980                 985                 990

tat gga ata ctc ttc ttg aag cac att gct gaa gaa ata gac gag aat   3024

Tyr Gly Ile Leu Phe Leu Lys His Ile Ala Glu Glu Ile Asp Glu Asn
        995                1000                1005

tct cca acc ttt tca gat ggt gtg ccc aag gaa gga ctt aga ata gaa   3072

Ser Pro Thr Phe Ser Asp Gly Val Pro Lys Glu Gly Leu Arg Ile Glu
   1010                1015                1020

gat gtt cta gtc aga att gct ctt ctg ata cta gtt cag gag aag gtg   3120

Asp Val Leu Val Arg Ile Ala Leu Leu Ile Leu Val Gln Glu Lys Val
```

1025 1030 1035 1040

aaa ttt gta gaa gat cat cca ggg aaa cct gtt ttc ccc tct cgc att 3168

Lys Phe Val Glu Asp His Pro Gly Lys Pro Val Phe Pro Ser Arg Ile
1045 1050 1055

ctt gaa aga ttc ccc gga ctg aga agt gga aaa att tgg aag gag gaa 3216

Leu Glu Arg Phe Pro Gly Leu Arg Ser Gly Lys Ile Trp Lys Glu Glu
1060 1065 1070

cat gac aag ata atg ata cgt gct gtt tta aag cat ggg tac gga cgg 3264

His Asp Lys Ile Met Ile Arg Ala Val Leu Lys His Gly Tyr Gly Arg
1075 1080 1085

tgg caa gct att gtt gat gac aaa gag ttg ggg atc caa gag ctt atc 3312

Trp Gln Ala Ile Val Asp Asp Lys Glu Leu Gly Ile Gln Glu Leu Ile
1090 1095 1100

tgc aaa gaa ttg aat ttc cct cac ata agt ttg tct gct gct gaa caa 3360

Cys Lys Glu Leu Asn Phe Pro His Ile Ser Leu Ser Ala Ala Glu Gln
1105 1110 1115 1120

gct ggt ttg cag ggg cag aat ggt agt ggg ggc tct aat ccg gga gca 3408

Ala Gly Leu Gln Gly Gln Asn Gly Ser Gly Gly Ser Asn Pro Gly Ala
1125 1130 1135

cag act aac cag aat cct gga agc gtt att act ggg aac aat aat gct 3456

Gln Thr Asn Gln Asn Pro Gly Ser Val Ile Thr Gly Asn Asn Asn Ala
1140 1145 1150

tct gct gat ggg gct caa gta aac tcg atg ttc tat tat cgg gac atg 3504

Ser Ala Asp Gly Ala Gln Val Asn Ser Met Phe Tyr Tyr Arg Asp Met
1155 1160 1165

cag aga cga ctt gtt gag ttt gtg aaa aag cga gtt ctg ctt ttg gag 3552

Gln Arg Arg Leu Val Glu Phe Val Lys Lys Arg Val Leu Leu Leu Glu
1170 1175 1180

aag gcg atg aat tat gaa tac gca gag gaa tat tat gga ctt ggt ggc 3600

Lys Ala Met Asn Tyr Glu Tyr Ala Glu Glu Tyr Tyr Gly Leu Gly Gly
1185 1190 1195 1200

```
tca tca tct atc cct act gaa gaa cca gaa gct gaa cca aag atc gct   3648

Ser Ser Ser Ile Pro Thr Glu Glu Pro Glu Ala Glu Pro Lys Ile Ala
                1205                1210                1215

gac aca gtg gga gtg agc ttt att gag gtt gat gat gaa atg ctt gat   3696

Asp Thr Val Gly Val Ser Phe Ile Glu Val Asp Asp Glu Met Leu Asp
            1220                1225                1230

gga ctt cct aag act gat cct atc act tca gaa gaa att atg ggg gct   3744

Gly Leu Pro Lys Thr Asp Pro Ile Thr Ser Glu Glu Ile Met Gly Ala
        1235                1240                1245

gct gtt gac aac aac caa gcg cgg gtc gaa ata gct caa cat tat aac   3792

Ala Val Asp Asn Asn Gln Ala Arg Val Glu Ile Ala Gln His Tyr Asn
    1250                1255                1260

cag atg tgc aaa ctt ctt gat gag aac gct cgg gaa tca gtc caa gca   3840

Gln Met Cys Lys Leu Leu Asp Glu Asn Ala Arg Glu Ser Val Gln Ala
1265                1270                1275                1280

tat gta aac aac caa cca ccg agt acc aag gtg aat gag agc ttc cgt   3888

Tyr Val Asn Asn Gln Pro Pro Ser Thr Lys Val Asn Glu Ser Phe Arg
                1285                1290                1295

gca ctc aaa tct atc aat ggt aac att aac aca atc ctt tcg att aca   3936

Ala Leu Lys Ser Ile Asn Gly Asn Ile Asn Thr Ile Leu Ser Ile Thr
            1300                1305                1310

tct gat caa tcc aag tca cat gaa gac gac acc aag cca gac cta aac   3984

Ser Asp Gln Ser Lys Ser His Glu Asp Asp Thr Lys Pro Asp Leu Asn
        1315                1320                1325

aat gtt gag atg aag gac acg gcc gaa gaa aca aaa ccg tta aga ggt   4032

Asn Val Glu Met Lys Asp Thr Ala Glu Glu Thr Lys Pro Leu Arg Gly
    1330                1335                1340

ggc gtc gtc gat ctg aat gtg gtg gag gga gag gag aac att gct gaa   4080

Gly Val Val Asp Leu Asn Val Val Glu Gly Glu Glu Asn Ile Ala Glu
1345                1350                1355                1360
```

gct agt gga agt gtt gat gta aaa atg gaa gaa gcc aaa gaa gaa gag 4128

Ala Ser Gly Ser Val Asp Val Lys Met Glu Glu Ala Lys Glu Glu Glu
1365 1370 1375

aag cca aag aac atg gtc gtt gat tgactcaact ggtaaatcaa gattc 4177

Lys Pro Lys Asn Met Val Val Asp
1380

<210> 2
<211> 1384
<212> PRT
<213> Arabidopsis thaliana

<400> 2

Met Ser Ser Leu Val Glu Arg Leu Arg Ile Arg Ser Asp Arg Lys Pro
1 5 10 15

Val Tyr Asn Leu Asp Asp Ser Asp Asp Asp Asp Phe Val Pro Lys Lys
20 25 30

Asp Arg Thr Phe Glu Gln Val Glu Ala Ile Val Arg Thr Asp Ala Lys
35 40 45

Glu Asn Ala Cys Gln Ala Cys Gly Glu Ser Thr Asn Leu Val Ser Cys
50 55 60

Asn Thr Cys Thr Tyr Ala Phe His Ala Lys Cys Leu Val Pro Pro Leu
65 70 75 80

Lys Asp Ala Ser Val Glu Asn Trp Arg Cys Pro Glu Cys Val Ser Pro
85 90 95

Leu Asn Glu Ile Asp Lys Ile Leu Asp Cys Glu Met Arg Pro Thr Lys
100 105 110

Ser Ser Glu Gln Gly Ser Ser Asp Ala Glu Pro Lys Pro Ile Phe Val
115 120 125

Lys Gln Tyr Leu Val Lys Trp Lys Gly Leu Ser Tyr Leu His Cys Ser
130 135 140

Trp Val Pro Glu Lys Glu Phe Gln Lys Ala Tyr Lys Ser Asn His Arg
145 150 155 160

Leu Lys Thr Arg Val Asn Asn Phe His Arg Gln Met Glu Ser Phe Asn
165 170 175

Asn Ser Glu Asp Asp Phe Val Ala Ile Arg Pro Glu Trp Thr Thr Val
180 185 190

Asp Arg Ile Leu Ala Cys Arg Glu Glu Asp Gly Glu Leu Glu Tyr Leu
195 200 205

Val Lys Tyr Lys Glu Leu Ser Tyr Asp Glu Cys Tyr Trp Glu Ser Glu
210 215 220

Ser Asp Ile Ser Thr Phe Gln Asn Glu Ile Gln Arg Phe Lys Asp Val
225 230 235 240

Asn Ser Arg Thr Arg Arg Ser Lys Asp Val Asp His Lys Arg Asn Pro
245 250 255

Arg Asp Phe Gln Gln Phe Asp His Thr Pro Glu Phe Leu Lys Gly Leu
260 265 270

Leu His Pro Tyr Gln Leu Glu Gly Leu Asn Phe Leu Arg Phe Ser Trp
275 280 285

Ser Lys Gln Thr His Val Ile Leu Ala Asp Glu Met Gly Leu Gly Lys
290 295 300

Thr Ile Gln Ser Ile Ala Leu Leu Ala Ser Leu Phe Glu Glu Asn Leu
305 310 315 320

Ile Pro His Leu Val Ile Ala Pro Leu Ser Thr Leu Arg Asn Trp Glu
325 330 335

Arg Glu Phe Ala Thr Trp Ala Pro Gln Met Asn Val Val Met Tyr Phe
340 345 350

Gly Thr Ala Gln Ala Arg Ala Val Ile Arg Glu His Glu Phe Tyr Leu
355 360 365

Ser Lys Asp Gln Lys Lys Ile Lys Lys Lys Lys Ser Gly Gln Ile Ser
370 375 380

Ser Glu Ser Lys Gln Lys Arg Ile Lys Phe Asp Val Leu Leu Thr Ser
385 390 395 400

Tyr Glu Met Ile Asn Leu Asp Ser Ala Val Leu Lys Pro Ile Lys Trp
405 410 415

Glu Cys Met Ile Val Asp Glu Gly His Arg Leu Lys Asn Lys Asp Ser
420 425 430

Lys Leu Phe Ser Ser Leu Thr Gln Tyr Ser Ser Asn His Arg Ile Leu
435 440 445

Leu Thr Gly Thr Pro Leu Gln Asn Asn Leu Asp Glu Leu Phe Met Leu
450 455 460

Met His Phe Leu Asp Ala Gly Lys Phe Gly Ser Leu Glu Glu Phe Gln
465 470 475 480

Glu Glu Phe Lys Asp Ile Asn Gln Glu Glu Gln Ile Ser Arg Leu His
485 490 495

Lys Met Leu Ala Pro His Leu Leu Arg Arg Val Lys Lys Asp Val Met
500 505 510

Lys Asp Met Pro Pro Lys Lys Glu Leu Ile Leu Arg Val Asp Leu Ser
515 520 525

Ser Leu Gln Lys Glu Tyr Tyr Lys Ala Ile Phe Thr Arg Asn Tyr Gln
530 535 540

Val Leu Thr Lys Lys Gly Gly Ala Gln Ile Ser Leu Asn Asn Ile Met
545 550 555 560

Met Glu Leu Arg Lys Val Cys Cys His Pro Tyr Met Leu Glu Gly Val
565 570 575

Glu Pro Val Ile His Asp Ala Asn Glu Ala Phe Lys Gln Leu Leu Glu
580 585 590

Ser Cys Gly Lys Leu Gln Leu Leu Asp Lys Met Met Val Lys Leu Lys
595 600 605

Glu Gln Gly His Arg Val Leu Ile Tyr Thr Gln Phe Gln His Met Leu
610 615 620

Asp Leu Leu Glu Asp Tyr Cys Thr His Lys Lys Trp Gln Tyr Glu Arg
625 630 635 640

Ile Asp Gly Lys Val Gly Gly Ala Glu Arg Gln Ile Arg Ile Asp Arg
645 650 655

Phe Asn Ala Lys Asn Ser Asn Lys Phe Cys Phe Leu Leu Ser Thr Arg
660 665 670

Ala Gly Gly Leu Gly Ile Asn Leu Ala Thr Ala Asp Thr Val Ile Ile
675 680 685

Tyr Asp Ser Asp Trp Asn Pro His Ala Asp Leu Gln Ala Met Ala Arg
690 695 700

Ala His Arg Leu Gly Gln Thr Asn Lys Val Met Ile Tyr Arg Leu Ile
705 710 715 720

Asn Arg Gly Thr Ile Glu Glu Arg Met Met Gln Leu Thr Lys Lys Lys
725 730 735

Met Val Leu Glu His Leu Val Val Gly Lys Leu Lys Thr Gln Asn Ile

740 745 750

Asn Gln Glu Glu Leu Asp Asp Ile Ile Arg Tyr Gly Ser Lys Glu Leu
755 760 765

Phe Ala Ser Glu Asp Asp Glu Ala Gly Lys Ser Gly Lys Ile His Tyr
770 775 780

Asp Asp Ala Ala Ile Asp Lys Leu Leu Asp Arg Asp Leu Val Glu Ala
785 790 795 800

Glu Glu Val Ser Val Asp Asp Glu Glu Glu Asn Gly Phe Leu Lys Ala
805 810 815

Phe Lys Val Ala Asn Phe Glu Tyr Ile Asp Glu Asn Glu Ala Ala Ala
820 825 830

Leu Glu Ala Gln Arg Val Ala Ala Glu Ser Lys Ser Ser Ala Gly Asn
835 840 845

Ser Asp Arg Ala Ser Tyr Trp Glu Glu Leu Leu Lys Asp Lys Phe Glu
850 855 860

Leu His Gln Ala Glu Glu Leu Asn Ala Leu Gly Lys Arg Lys Arg Ser
865 870 875 880

Arg Lys Gln Leu Val Ser Ile Glu Glu Asp Asp Leu Ala Gly Leu Glu
885 890 895

Asp Val Ser Ser Asp Gly Asp Glu Ser Tyr Glu Ala Glu Ser Thr Asp
900 905 910

Gly Glu Ala Ala Gly Gln Gly Val Gln Thr Gly Arg Arg Pro Tyr Arg
915 920 925

Arg Lys Gly Arg Asp Asn Leu Glu Pro Thr Pro Leu Met Glu Gly Glu
930 935 940

```
Gly Arg Ser Phe Arg Val Leu Gly Phe Asn Gln Ser Gln Arg Ala Ile
945                 950                 955                 960

Phe Val Gln Thr Leu Met Arg Tyr Gly Ala Gly Asn Phe Asp Trp Lys
                965                 970                 975

Glu Phe Val Pro Arg Leu Lys Gln Lys Thr Phe Glu Glu Ile Asn Glu
            980                 985                 990

Tyr Gly Ile Leu Phe Leu Lys His  Ile Ala Glu Glu Ile  Asp Glu Asn
        995                 1000                1005

Ser Pro  Thr Phe Ser Asp Gly  Val Pro Lys Glu Gly  Leu Arg Ile
    1010                1015                1020

Glu Asp  Val Leu Val Arg Ile  Ala Leu Leu Ile Leu  Val Gln Glu
    1025                1030                1035

Lys Val  Lys Phe Val Glu Asp  His Pro Gly Lys Pro  Val Phe Pro
    1040                1045                1050

Ser Arg  Ile Leu Glu Arg Phe  Pro Gly Leu Arg Ser  Gly Lys Ile
    1055                1060                1065

Trp Lys  Glu Glu His Asp Lys  Ile Met Ile Arg Ala  Val Leu Lys
    1070                1075                1080

His Gly  Tyr Gly Arg Trp Gln  Ala Ile Val Asp Asp  Lys Glu Leu
    1085                1090                1095

Gly Ile  Gln Glu Leu Ile Cys  Lys Glu Leu Asn Phe  Pro His Ile
    1100                1105                1110

Ser Leu  Ser Ala Ala Glu Gln  Ala Gly Leu Gln Gly  Gln Asn Gly
    1115                1120                1125

Ser Gly  Gly Ser Asn Pro Gly  Ala Gln Thr Asn Gln  Asn Pro Gly
    1130                1135                1140
```

Ser Val Ile Thr Gly Asn Asn Asn Ala Ser Ala Asp Gly Ala Gln
1145 1150 1155

Val Asn Ser Met Phe Tyr Tyr Arg Asp Met Gln Arg Arg Leu Val
1160 1165 1170

Glu Phe Val Lys Lys Arg Val Leu Leu Leu Glu Lys Ala Met Asn
1175 1180 1185

Tyr Glu Tyr Ala Glu Glu Tyr Tyr Gly Leu Gly Gly Ser Ser Ser
1190 1195 1200

Ile Pro Thr Glu Glu Pro Glu Ala Glu Pro Lys Ile Ala Asp Thr
1205 1210 1215

Val Gly Val Ser Phe Ile Glu Val Asp Asp Glu Met Leu Asp Gly
1220 1225 1230

Leu Pro Lys Thr Asp Pro Ile Thr Ser Glu Glu Ile Met Gly Ala
1235 1240 1245

Ala Val Asp Asn Asn Gln Ala Arg Val Glu Ile Ala Gln His Tyr
1250 1255 1260

Asn Gln Met Cys Lys Leu Leu Asp Glu Asn Ala Arg Glu Ser Val
1265 1270 1275

Gln Ala Tyr Val Asn Asn Gln Pro Pro Ser Thr Lys Val Asn Glu
1280 1285 1290

Ser Phe Arg Ala Leu Lys Ser Ile Asn Gly Asn Ile Asn Thr Ile
1295 1300 1305

Leu Ser Ile Thr Ser Asp Gln Ser Lys Ser His Glu Asp Asp Thr
1310 1315 1320

Lys Pro Asp Leu Asn Asn Val Glu Met Lys Asp Thr Ala Glu Glu
1325 1330 1335

Thr Lys Pro Leu Arg Gly Gly Val Val Asp Leu Asn Val Val Glu
1340 1345 1350

Gly Glu Glu Asn Ile Ala Glu Ala Ser Gly Ser Val Asp Val Lys
1355 1360 1365

Met Glu Glu Ala Lys Glu Glu Glu Lys Pro Lys Asn Met Val Val
1370 1375 1380

Asp

<210> 3

<211> 19

<212> DNA

<213> Artificial Sequence

<220>

<221> misc_feature

<222> 18-19

<223> AFLP Primer EcoRI for AFLP Mapping Analysis in Example 1; n may be a, g, c or t

<400> 3

agactgcgta ccatttcnn 19

<210> 4

<211> 19

<212> DNA

<213> Artificial Sequence

<220>

<221> misc_feature

<222> 17-19

<223> AFLP Primer MseI for AFLP Mapping Analysis in Example 1;
n may be a, g, c or t

<400> 4

gatgagtcct gagtaannn 19

<210> 5

<211> 21

<212> DNA

<213> Arabidopsis thaliana

<220>

<223> Primers for PCR of Example 2;
sequence complementary to nucleotides 1725-1745 of SEQ ID NO:1

<400> 5

tgttgagcca gttattcacg a 21

<210> 6

<211> 21

<212> DNA

<213> Arabidopsis thaliana

<220>

<223> Primers for PCR of Example 2;
sequence complementary to nucleotides 1934-1914 in SEQ ID NO:1

<400> 6

acctttccat caattcgctc g 21

<210> 7
<211> 30

<212> DNA

<213> Artificial Sequence

<220>

<221> misc_feature

<222> 1-30

<223> Primers for PCR of Example 2

<400> 7

ccgctcgaga accccaatga ccagctcagt 30

<210> 8

<211> 21

<212> DNA

<213> Artificial Sequence

<220>

<221> misc_feature

<222> 1-21

<223> Primers for PCR of Example 2;
sequence complementary to nucleotides 672-652 of
LEC1 cDNA sequence

<400> 8

ccttcttcac ttatactgac c 21

<210> 9

<211> 21

<212> DNA

<213> Arabidopsis thaliana

<220>

<223> Primers for PCR of Example 2;
nucleotides 65-85 of ROC3 cDNA sequence

<400> 9

aagtctactt cgacatgacc g 21

<210> 10

<211> 21

<212> DNA

<213> Arabidopsis thaliana

<220>

<223> Primers for PCR of Example 2;
sequence complementary to nucleotides 524-504 of ROC3
cDNA sequence

<400> 10

cttccagagt cagatccaac c 21

<210> 11

<211> 30

<212> DNA

<213> Arabidopsis thaliana

<220>

<223> Primers for PCR of Example 4;
represent nucleotides 895-924 in SEQ ID NO:1 wherein nucleotide
907 is changed from "a" to "g"

<400> 11

gaaatgggac taggcaggac aattcaaagc 30

<210> 12

<211> 30

<212> DNA

<213> Arabidopsis thaliana

<220>

<223> Primers for PCR of Example 4;
represent sequence complementary to nucleotides
924-895 in SEQ ID NO:1, with nucleotide 911 changed from "t" to "c".

<400> 12

gctttgaatt gtcctgccta gtcccatttc 30

<210> 13

<211> 47

<212> DNA

<213> Artificial Sequence

<220>

<221> misc_feature

<222> 1-47

<223> Primers for PCR of Example 4

<400> 13

aagccaaaga acatggtcgt tgatctagag gatcctgaag ctcgaaa 47

<210> 14

<211> 52

<212> DNA

<213> Artificial Sequence

<220>

<221> misc_feature

<222> 1-52

<223> Primers for PCR of Example 4

<400> 14

gaatcttgat ttaccagttg agtcattttt gatgaaacag aagctttttg at 52

<210> 15

<211> 21

<212> DNA

<213> Arabidopsis thaliana

<220>

<223> Primers for PCR of Example 4;
represent sequence complementary to nucleotides 4152-4132 in SEQ ID NO:1

<400> 15

atcaacgacc atgttctttg g 21

<210> 16

<211> 22

<212> DNA

<213> Arabidopsis thaliana

<220>

<223> Primers for PCR of Example 4;
represent nucleotides 4153-4174 in SEQ ID NO:1

<400> 16

tgactcaact ggtaaatcaa ga 22

<210> 17

<211> 30

<212> DNA

<213> Artificial Sequence

<220>

<221> misc_feature

<222> 1-30

<223> Primers for PCR of Example 5

<400> 17

ccgctcgagt gagtagtttg gtggagaggc 30

<210> 18

<211> 30

<212> DNA

<213> Artificial Sequence

<220>

<221> misc_feature

<222> 1-30

<223> Primers for PCR of Example 5

<400> 18

ccggaattcc atcggaggaa ccttgttcac 30

<210> 19

<211> 30

<212> DNA

<213> Artificial Sequence

<220>

<221> misc_feature

<222> 1-30

<223> Primers for PCR of Example 5

<400> 19

cgcggatccc atcggaggaa ccttgttcac 30

<210> 20

<211> 30

<212> DNA

<213> Artificial Sequence

<220>

<221> misc_feature

<222> 1-30

<223> Primers for PCR of Example 5

<400> 20

tgctctagat gagtagtttg gtggagaggc 30

<210> 21

<211> 30

<212> DNA

<213> Artificial Sequence

<220>

<221> misc_feature

<222> 1-30

<223> Primers for PCR of Example 5

<400> 21

ccgctcgagc cctcacataa gtttgtctgc 30

<210> 22

<211> 30

<212> DNA

<213> Artificial Sequence

<220>

<221> misc_feature

<222> 1-30

<223> Primers for PCR of Example 5

<400> 22

ccggaattcg tcttaggaag tccatcaagc 30

<210> 23

<211> 30

<212> DNA

<213> Artificial Sequence

<220>

<221> misc_feature

<222> 1-30

<223> Primers for PCR of Example 5

<400> 23

cgcggatccg tcttaggaag tccatcaagc 30

<210> 24

<211> 30

<212> DNA

<213> Artificial Sequence

<220>

<221> misc_feature

<222> 1-30

<223> Primers for PCR of Example 5

<400> 24

tgctctagac cctcacataa gtttgtctgc 30

<210> 25

<211> 31

<212> DNA

<213> Artificial Sequence

<220>

<221> misc_feature

<222> 1-31

<223> Primers for PCR of Example 6

<400> 25

cgcggatcct ttttccactt ctcagtccgg g 31

<210> 26

<211> 34

<212> DNA

<213> Artificial Sequence

<220>

<221> misc_feature

<222> 1-34

<223> Primers for PCR of Example 4

<400> 26

cttcgaactc gagggatccc catggctagc agct 34

<210> 27

<211> 34

<212> DNA

<213> Artificial Sequence

<220>

<221> misc_feature

<222> 1-34

<223> Primers for PCR of Example 4

<400> 27

gctagccatg gggatccctc gagttcgaag gtac 34

<210> 28

<211> 12

<212> DNA

<213> Artificial Sequence

<220>

<221> misc_feature

<222> 1-12

<223> Primers for forming cassette inserted into pCAMBIA3300 in Example 4

<400> 28

ccaggtacct gg 12

<210> 29

<211> 20

<212> DNA

<213> Artificial Sequence

<220>

<221> misc_feature

<222> 1-20

<223> Primers for forming cassette inserted into pCAMBIA3300 in Example 4

<400> 29

aattccaggt acctggcatg 20

<210> 30

<211> 38

<212> DNA

<213> Artificial Sequence

<220>

<221> misc_feature

<222> 1-38

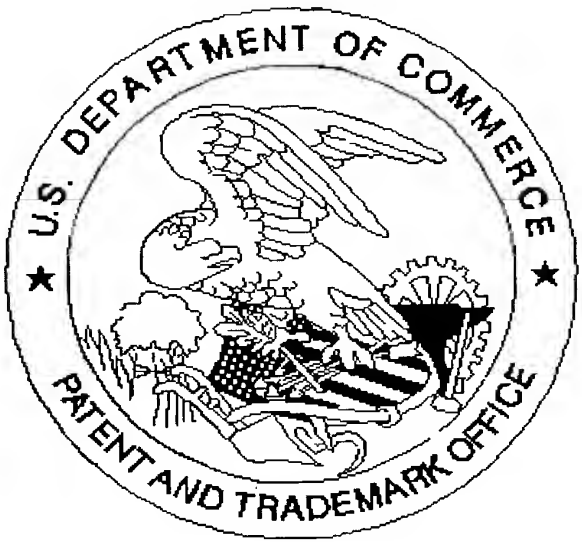
<223> Sequence for forming clone of the rat glucocorticoid receptor in Example 4

<400> 30

tctagaggat cctgaagctc gaaaaacaaa gaaaaaaa 38

United States Patent & Trademark Office

Office of Initial Patent Examination -- Scanning Division



Application deficiencies found during scanning:

☐ Page(s)_____of_____ were not present for scanning. (Document title)

☐ Page(s)_____of_____were not present for scanning. (Document title)

☐ ***Scanned copy is best available.***

Some hauky figures are too dark.